BIOPRESS DIVISION
PRESENTS

CONTEMPORARY DRUG DESIGN & THERAPEUTIC INNOVATION

EDITED BY
DR RICHA SHARMA

CO-EDITOR
DR K.P. BEENA

PUBLISHER

BIOPRESS

Title - Contemporary Drug Design & Therapeutic Innovation

Copyright - ©2026 Swalife Foundation

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission from the publisher, except for brief quotations used in critical reviews and other non-commercial uses.

Published by: Swalife Foundation

Publishing Under Biopress Division of Swalife Foundation

Address – Swalife Foundation JSPM college road

Pune, India – 411041

Email: swalifefoundation@gmail.com
Printed by – Swalife foundation, Pune, India.

Editor – Dr Richa Sharma

Co-Editor - Dr K. P. Beena

ISBN: 978-93-49682-74-0

Contributing Authors - This book features chapter contributions from experts in the field of Pharmacy & Biotechnology, as follows**:**

Dr Richa Sharma, Gajanand Sharma, H.M Rakshith, Jagruti Gosavi, Rajeshwari Dughde, Trupti Kadam, Ashwini Badhe, Dr K. P. Beena, Arghya Saha Chowdhury, Ilias Ahmed, Paragmoni Bora, Dr C. Jaganmohan, Dr Srinidhi. S., Dr.S.Sasikala, E. Nivetha, R. Geerhanjali.

Disclaimer - This *book contains information intended for educational purposes and does not constitute professional advice. Readers are encouraged to consult professionals for specific guidance.*

Content Responsibility Policy - *The content of this book is the sole responsibility of the respective authors and does not reflect the views of the editor and publisher. Each author ensures their work is original and compliant with ethical and copyright standards. The publisher is not liable for any errors or claims arising from individual chapters.*

BIOPRESS

Preface

The field of drug discovery and therapeutic innovation is rapidly evolving through advancements in artificial intelligence, computational biology, molecular modeling, precision medicine, and translational research. Modern pharmaceutical research now combines computational and experimental approaches to accelerate the development of safer and more effective therapies.

Contemporary Drug Design & Therapeutic Innovation has been developed to provide readers with updated knowledge on emerging strategies and technologies in modern drug discovery. The book covers important topics including AI-assisted drug delivery, structure-based drug design, molecular docking, lead optimization, ADMET prediction, computational screening, precision therapeutics, translational cancer research, and regulatory aspects of drug development.

This book is intended for students, researchers, academicians, and professionals in pharmacy, biotechnology, life sciences, and healthcare fields. The editors sincerely appreciate the valuable contributions of all authors and thank Swalife Foundation and the Biopress Division for their support in publishing this book.

We hope this book serves as a useful academic and research resource and encourages further innovation in therapeutic science and drug discovery.

Dr. Richa Sharma
Editor
Professor, Department of Microbiology,
Mahatma Gandhi University of Medical Sciences & Technology, Jaipur, India

Dr. K. P. Beena
Co-Editor
Associate Professor, Department Of Pharmaceutical Chemistry,
Sri Ramakrishna Institute of Paramedical Sciences, College Of Pharmacy,
Coimbatore, Tamil Nadu, India

BIOPRESS

BIOPRESS

Index

1. **Chapter 1: Artificial Intelligence assisted Drug Delivery: Validation & Identification**

   Author: ***Richa Sharma, Gajanand Sharma, H.M Rakshith***

   Page No. 1-19

2. **Chapter 2: Structure-Based Drug Design and Molecular Docking Approaches in Contemporary Therapeutic Innovation**

   Author: ***Jagruti Gosavi, Rajeshwari Dughde, Trupti Kadam***

   Page No. 20-34

3. **Chapter 3: Lead Optimization Strategies in Small Molecule Development**

   Author: ***Ashwini Badhe***

   Page No.35-50

4. **Chapter 4: ADMET Prediction and Pharmacokinetic Modeling**

   *Author****: Dr K. P. Beena***

   Page No. 51-68

5. **Chapter 5: Drug Repurposing and Computational Screening Approaches**

   Author: ***Arghya Saha Chowdhury, Ilias Ahmed, Paragmoni Bora***

   Page No. 69-95

6. **Chapter 6: Biomarkers in Precision Therapeutics**

   Author: ***Dr C. Jaganmohan, Dr Srinidhi. S***

   Page No. 96-110

7. **Chapter 7: Translational Approaches in Cancer Research: Therapeutic Innovation and Preclinical Evaluation Using Indian Normal and Cancer Cell Line Models**

   Author: ***Dr.S.Sasikala, E. Nivetha, R. Geerhanjali***

   Page No. 111-127

8. **Chapter 8: Regulatory Considerations in New Drug Approval Pathways**

   Author: ***Ashwini Badhe***

   Page No. 128- 145

BIOPRESS

BIOPRESS

Chapter 1

Artificial Intelligence assisted Drug Delivery: Validation & Identification

Richa Sharma*[1] Gajanand Sharma[2] H.M Rakshith[3]

[1]Department of Microbiology, Mahatma Gandhi University of Medical Sciences & Technology, Jaipur.

[2]Department of Chemistry, MPS International, Jaipur.

[3]Department of Clinical Psychology, Mahatma Gandhi University of Medical Science & Technology, Jaipur

---

Abstract:

Artificial Intelligence (AI) has been an emerging tool in the transformation for drug discovery, and it has revolutionised the way researchers identify and develop new therapeutic compounds. AI has emerged as a powerful tool offering innovative solutions by easily extracting the complex biological data and modelling molecular interactions in a precise and accurate manner. By leveraging AI algorithms, researchers can analyse vast amounts of biological data, predict drug-target interactions, and design novel molecules with enhanced specificity and efficacy. AI-driven approaches such as machine learning, deep learning, and natural language processing have enabled the discovery of new drug candidates in a fraction of the time and cost compared to traditional methods. Traditionally, many processes are used for target identification, and now these processes have evolved to include advanced methods such as AI-driven multi-centric approaches. AI aids in analysing massive databases (DNA, RNA, protein, and metabolites) to discover hidden patterns, which are considered a basic need in the current 21st-century scenario. This chapter highlights key AI techniques and applications in drug design, discovery, validation & identification, AI in clinical trials, challenges, limitations & benefits of AI in drug discovery.

Keywords: *Artificial Intelligence, Drug design, validation, target identification.*

1 Introduction: Drug design

Drug design—often referred to as rational drug design—is the inventive process of finding new medications based on the knowledge of a biological target. It involves identifying a specific molecule in the body—usually a protein—linked to a disease and then engineering a "key" (the drug) that fits perfectly into that protein's "lock" to turn a biological process on or off. drug design is an interdisciplinary feat. It blends **pharmacology**, **biochemistry**, and **computer science.** Drug design is the inventive process of finding new medications based on the knowledge of a biological target. Drug development and discovery includes preclinical

research on cell-based and animal models and clinical trials on humans, and finally move forward to the step of obtaining regulatory approval in order to market the drug. Modern drug discovery involves the identification of screening hits, medicinal chemistry and optimization of those hits to increase the affinity, selectivity (to reduce the potential of side effects), efficacy/potency, metabolic stability (to increase the half-life), and oral bioavailability. Once a compound that fulfills all of these requirements has been identified, it will begin the process of drug development prior to clinical trials (Zhou & Zhang, 2017). The introduction of computational approaches has drastically changed the landscape of drug development. By employing molecular modeling, docking, and dynamics simulations, researchers can now visualize and predict drug-target interactions at the atomic level. These advancements have allowed the transition from random screening to rational drug design, where compounds are engineered with specific properties in mind. Moreover, with the rise of artificial intelligence (AI) and big data, the integration of patient-specific genetic, proteomic, and metabolic data has facilitated the advent of personalized medicine (Zende et al., 2025).

1.1 Brief Overview of AI in Drug Discovery:

In 2026, **Artificial Intelligence (AI)** has transitioned from a experimental tool to the foundational driver of pharmaceutical innovation. The industry is moving away from traditional trial-and-error toward a "computation-first" model, where billions of molecules are screened virtually before a single test tube is touched.

Over the past few years, there has been a drastic increase in data digitalization in the pharmaceutical sector. However, this digitalization comes with the challenge of acquiring, scrutinizing, and applying that knowledge to solve complex clinical problems (Ramesh et al., 2004). This motivates the use of AI, because it can handle large volumes of data with enhanced automation (Miles & Walker, 2006). AI is a technology-based system involving various advanced tools and networks that can mimic human intelligence (Yang & Siau, 2018; Wirtz, et al., 2018). AI utilizes systems and software that can interpret and learn from the input data to make independent decisions for accomplishing specific objectives. Its applications are continuously being extended in the pharmaceutical field. The impact of AI is particularly profound in the healthcare and pharmaceutical sectors. AI-powered technologies are being leveraged to address critical challenges in drug discovery and development, improve patient care, and enhance healthcare delivery systems (Mak et al., 2024; Vora et al., 2023). AI refers to a wide array of technologies that allow machines to replicate human intelligence, including learning, reasoning, and decision-making (Fetzer, 1990). ML, a branch of AI, enables systems to learn from data and enhances their performance on specific tasks through experience, without the need for explicit programming (Alpaydin, 2021). DL is a specialized type of ML that uses Artificial Neural Networks (ANN) with multiple layers ("deep" networks) to identify complex patterns and derive meaningful insights from data (Goodfellow, 2016). Neural networks (NN), inspired by the human brain, are interconnected networks of nodes that process information in a layered fashion, enabling them to learn and make predictions (Gurney, 2018).

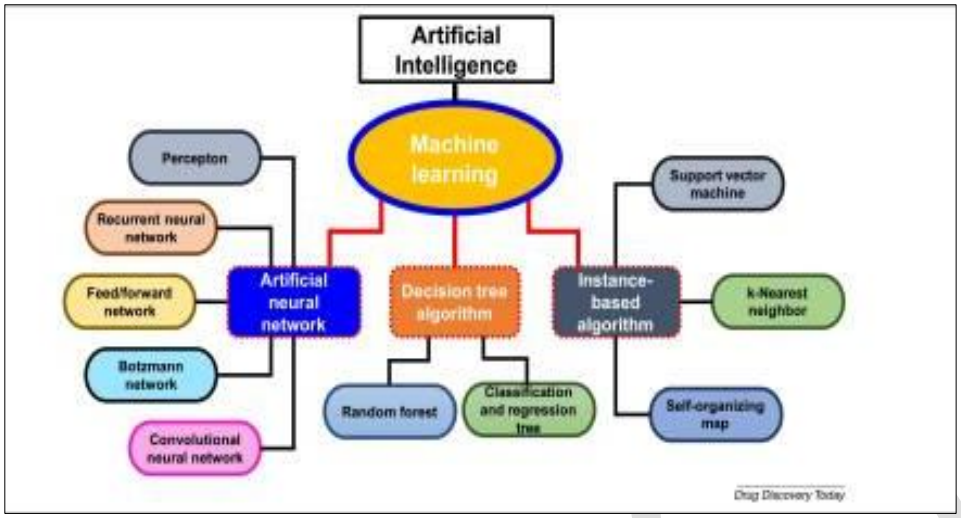


Figure 1.1: Method domains of artificial intelligence (AI). This figure shows different AI method domains along with their subfields that can be implemented in different fields drug discovery and development.

Key Pillars of AI in Drug discovery

1. **Target Identification :** AI platforms analyze vast "omics" datasets—genomics, proteomics, and transcriptomics—to find hidden biological drivers of diseases like oncology and rare genetic disorders.
2. **De Novo Molecular Design :** Using **Generative AI**, researchers can now "print" entirely new chemical structures. Instead of searching through existing libraries, AI models like **RNNs** (Recurrent Neural Networks) and **Transformers** design molecules from scratch that are optimized for potency and safety simultaneously.
3. **Structure Prediction :** The release and evolution of **AlphaFold 3** has been a watershed moment. It can now predict the 3D structures and interactions of not just proteins, but also DNA, RNA, and ligands with over 50% more accuracy than previous methods.

1.2 In AI-based drug design, validation and identification are crucial steps.

Validation:
Molecular docking: Validating predicted binding affinities and poses

What is Molecular Docking?
Molecular docking is a computational method used to predict the binding affinity and pose of a ligand (small molecule) to a protein (receptor). It's a crucial step in structure-based drug design.
Software Options: (Morris et al., 2009)

1. AutoDock: A popular open-source software for molecular docking. It uses a genetic algorithm to predict binding poses and affinities.
2. DOCK: A widely used software for molecular docking, developed at UCSF. It uses a shape-based approach to predict binding poses.

3. Glide: A commercial software for molecular docking, developed by Schrödinger. It uses a hierarchical approach to predict binding poses and affinities.

Key Factors Affecting Docking Accuracy:

1. Protein structure preparation: Ensuring the protein structure is accurate and complete.
2. Ligand preparation: Ensuring the ligand structure is accurate and protonated correctly.
3. Docking algorithm: Choosing the right algorithm (e.g., genetic algorithm, shape-based).
4. Scoring function: Evaluating binding affinity (e.g., force field-based, empirical).

Improving Docking Results:

1. Use multiple docking software: Compare results from different software to increase confidence.
2. Use protein flexibility: Account for protein flexibility using methods like induced fit docking.
3. Use water molecules: Include explicit water molecules to improve accuracy.
4. Validate results: Use experimental data (e.g., X-ray crystallography, NMR) to validate docking results.

Molecular dynamics simulations: Assessing stability and dynamics of protein-ligand complexes

What are Molecular Dynamics Simulations?
Molecular dynamics (MD) simulations are computational methods used to study the dynamics and stability of protein-ligand complexes over time. They provide insights into protein-ligand interactions, conformational changes, and binding free energies.

Software : GROMACS, AMBER, NAMD (Abraham et al., 2015)

1. GROMACS: A popular open-source software for MD simulations, known for its high performance and versatility.
2. AMBER: A widely used software for MD simulations, particularly suited for biomolecular systems.
3. NAMD: A scalable software for MD simulations, developed at UIUC, suitable for large systems.

Key Steps in MD Simulations:

1. System preparation: Preparing the protein-ligand complex, including solvation and ionization.
2. Energy minimization: Minimizing the system's energy to remove clashes.
3. Equilibration: Equilibrating the system to reach a stable state.
4. Production run: Running the simulation to collect data.

BIOPRESS

Analysis and Interpretation:

1. RMSD (Root Mean Square Deviation): Assessing protein-ligand stability.
2. RMSF (Root Mean Square Fluctuation): Evaluating protein flexibility.
3. Hydrogen bonding: Analyzing protein-ligand interactions.
4. Free energy calculations: Estimating binding free energies using methods like MM-PBSA or MM-GBSA.

Improving MD Simulations:

1. Use a suitable force field: Choose a force field suitable for your system (e.g., AMBER, CHARMM).
2. Use a sufficient simulation time: Ensure the simulation is long enough to capture relevant dynamics.
3. Use multiple replicas: Run multiple simulations to improve statistical accuracy.
4. Validate results: Compare simulation results with experimental data.

Free energy calculations: Estimating binding free energies

Free energy calculations estimate the binding free energy (ΔG) of a protein-ligand complex, predicting the strength of binding.

Methods: MM-PBSA, MM-GBSA, FEP (Wang et al, 2015).

1. MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area): Estimates ΔG using molecular mechanics energies and implicit solvation.
2. MM-GBSA (Molecular Mechanics Generalized Born Surface Area): Similar to MM-PBSA, but uses a more approximate solvation model.
3. FEP (Free Energy Perturbation): Estimates ΔG by perturbing the ligand's interactions with the protein.

Key Aspects:

1. Accuracy: MM-PBSA and MM-GBSA are widely used, but may have limitations in accuracy. FEP can be more accurate but is computationally demanding.
2. Computational cost: MM-PBSA and MM-GBSA are relatively fast, while FEP requires extensive sampling.
3. System preparation: Proper system preparation is crucial for accurate results.

BIOPRESS

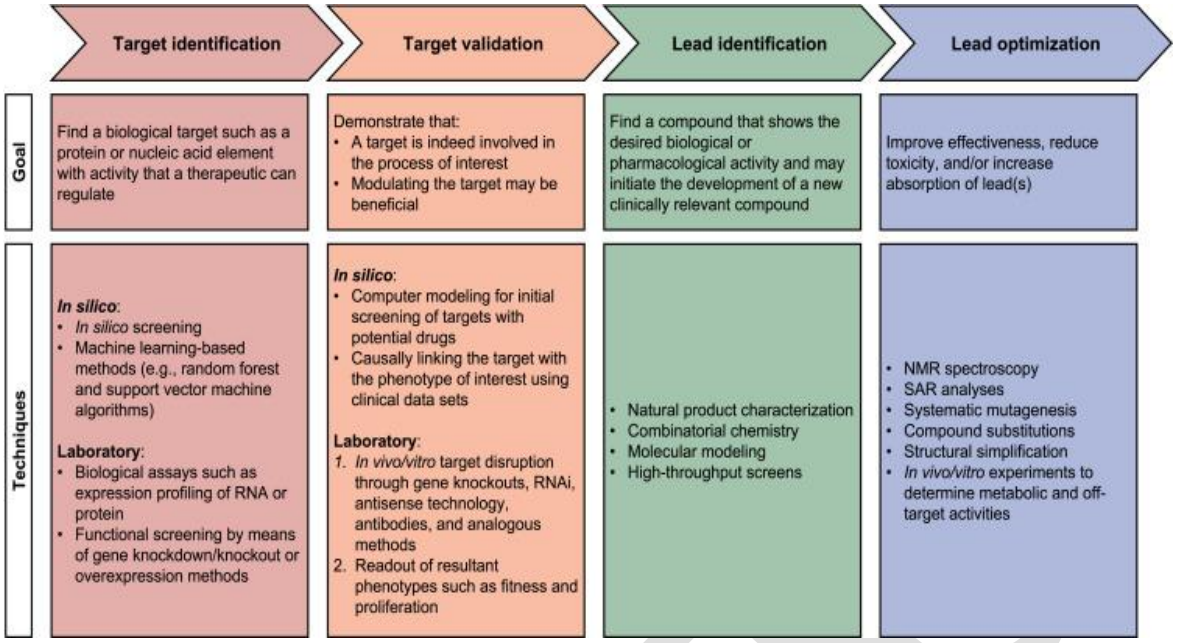

Target identification
Target validation
Lead identification
Lead optimization
Goal
Find a biological target such as a protein or nucleic acid element with activity that a therapeutic can regulate
Demonstrate that:
• A target is indeed involved in the process of interest
• Modulating the target may be beneficial
Find a compound that shows the desired biological or pharmacological activity and may initiate the development of a new clinically relevant compound
Improve effectiveness, reduce toxicity, and/or increase absorption of lead(s)
Techniques
In silico:
• In silico screening
• Machine learning-based methods (e.g., random forest and support vector machine algorithms)
Laboratory:
• Biological assays such as expression profiling of RNA or protein
• Functional screening by means of gene knockdown/knockout or overexpression methods
In silico:
• Computer modeling for initial screening of targets with potential drugs
• Causally linking the target with the phenotype of interest using clinical data sets
Laboratory:
1. In vivo/vitro target disruption through gene knockouts, RNAi, antisense technology, antibodies, and analogous methods
2. Readout of resultant phenotypes such as fitness and proliferation
• Natural product characterization
• Combinatorial chemistry
• Molecular modeling
• High-throughput screens
• NMR spectroscopy
• SAR analyses
• Systematic mutagenesis
• Compound substitutions
• Structural simplification
• In vivo/vitro experiments to determine metabolic and off-target activities

Figure 1.3: Target Identification

1.3 Identification:

1. Ligand-based approaches: Identifying potential leads using known active compounds Methods: Pharmacophore modeling, QSAR (Tropsha, 2010).
2. Structure-based approaches: Identifying potential leads using protein structure information Methods: Virtual screening, molecular docking (Kitchen et al., 2004)
3. De novo design: Generating novel molecules with desired properties
   Methods: Generative models, reinforcement learning (Meyers et al., 2021)

José Jiménez-Luna et al. investigated the use of AI-assisted computational models in drug discovery, specifically in ligand-based quantitative structure-activity and property (QSAR/QSPR) and De Novo drug design. Successful progress has been made in predicting pharmacokinetic parameters and biological activity using QSAR/QSPR. Molecular descriptors are used to provide machine-readable numbers representing the structural characteristics of molecules, including functional groups, pharmacophore distribution, and physicochemical properties (Lim et al, 2018). Although ligand and/or structure-based approaches are the most well-known methods for creating unique molecular profiles with efficient pharmacological potency and characteristics, computer-assisted de novo drug creation remains a challenging task (Sarella, et al., 2024).

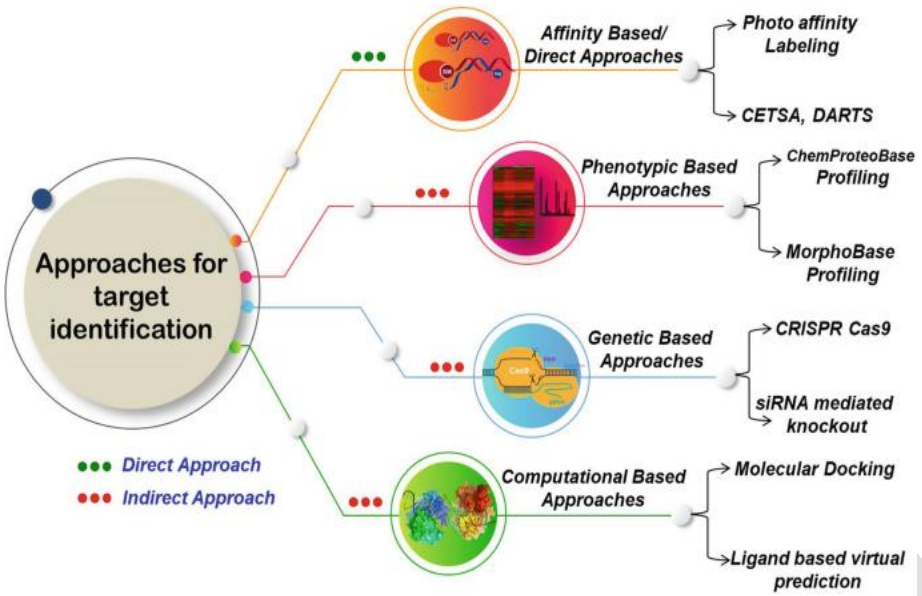


Figure 1.4: Target Identification Approaches in Drug Discovery

Some popular validation metrics:

1. Docking score: Evaluating binding affinity. -Higher scores typically indicate stronger binding. -Dependent on the scoring function used.
2. RMSD (Root Mean Square Deviation): Assessing pose accuracy. -Measures deviation between predicted and experimental ligand poses. -RMSD < 2 Å typically indicates good pose prediction.
3. AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Evaluating model performance (Hanley et al., 1982). -Measures ability to distinguish actives from decoys - AUC-ROC = 1 indicates perfect separation, 0.5 indicates random performance.

Other Metrics:

1. Enrichment Factor (EF): Evaluates early enrichment of actives.
2. BEDROC (Boltzmann-Enhanced Discrimination of ROC): Evaluates early enrichment with a focus on early recognition.
3. R^2: Evaluates correlation between predicted and experimental binding affinities.

1.4 Benefits of AI in Drug development

- Rapid screening of large chemical libraries: AI algorithms can quickly sort through millions of molecules to identify the most promising ones for further investigation.
- Improved prediction accuracy for drug efficacy: AI algorithms can accurately predict the efficacy of potential compounds, allowing researchers to focus on the most promising candidates.
- Early detection of potential side effects: AI systems can predict the possible side effects of a compound before human testing, reducing the risk of adverse outcomes.

- Personalized medicine: AI can help researchers develop drugs tailored to specific individuals based on their genetic composition and other health information.
- Lower drug development costs: By optimizing the drug development process, AI drug discovery can significantly reduce the time and cost required to bring new drugs to market.
- AI is also being used to identify drug interactions, enhance drug efficacy, and optimize clinical trials. These applications of AI in drug development are opening up new avenues for research and have the potential to revolutionize the pharmaceutical industry

1.5 Conventional versus AI assisted drug design

The pharmaceutical landscape in 2026 is defined by a hybrid model where AI handles the "heavy lifting" of data and design, while conventional biology focuses on the high-stakes physical validation.

Table1: Differentiation between Conventional versus AI assisted Drug Design

Metric	Conventional Design	AI-Assisted Design (2026)
Discovery Timeline	5–7 Years	1–2 Years
Synthesis Phase	2,000+ compounds synthesized	<200 highly-optimized compounds
Average Cost	~$2.6 Billion (total R&D)	~40-50% reduction in early-stage cost
Attrition Rate	90% failure in clinical trials	Predicted 20-30% improvement in trial success

1.6 AI Methodologies & Algorithmic Designs

AI-enhanced High-Throughput Screening (HTS) increasingly integrates neural networks, often employing CNN or RNN-based layers for rapid data extraction and analysis from high-content imaging assays. Furthermore, contemporary AI-driven HTS methodologies are leveraging transfer learning to effectively incorporate the inherent multifidelity of HTS data. Traditional HTS funnels generate data across tiers, from large-scale, lower-fidelity primary screens to smaller, high-fidelity confirmatory assays. Recent research, as exemplified by (Buterez et al., 2023) underscores the efficacy of transfer learning, particularly with Graph Neural Networks (GNNs), in enhancing molecular property prediction through the strategic integration of these multifidelity HTS data modalities. This approach acknowledges that primary screening data, while noisier, encompasses a vast chemical space, offering a valuable, inexpensive proxy to guide predictions on sparse, high-fidelity confirmatory screen data. Data sets like MF-PCBA, introduced by (Buterez et al., 2024) to benchmark such methods, provide a publicly

available collection of standardized multifidelity HTS data, facilitating the development and evaluation of AI/ML models that effectively utilize multitiered experimental information.

Virtual screening utilizes diverse AI algorithms, including deep learning models and GNNs, to prioritize candidate molecules based on predicted properties. MSGNN-DTA, for example, combines atom-, motif-, and protein-level graphs with gated skip-connections for robust binding affinity prediction, achieving a low RMSE of 1.237 on the KIBA benchmark and demonstrating practical utility in virtual screening acceleration through an FDA-approved drug case study.

AI-driven drug repurposing utilizes various machine learning methodologies, often employing supervised learning models like Random Forest or Support Vector Machines (SVMs) trained on drug activity data. More advanced approaches, like meta-learning frameworks such as Meta-GAT, are being explored to address data scarcity in repurposing by leveraging transfer learning methodologies across diverse drug and disease contexts, aiming to improve prediction accuracy, especially in low-data regimes (Ye, et al., 2021). Large Language Models (LLMs) are also emerging as promising tools for drug repurposing. Frameworks like DrugReAlign, developed by (Wei et al., 2024) demonstrate that LLMs, when combined with multisource prompting, can effectively utilize vast knowledge bases to identify potential drug repurposing opportunities and mitigate challenges such as 'hallucinations' through incorporating multisource information and spatial interaction data.

1.7 .Opportunities of AI in drug discovery

Traditional drug discovery methods have historically been expensive, time-consuming, and error-prone. However, the emergence of machine learning drug discovery is transforming the industry Researchers can analyze vast amounts of data and identify potential chemical compounds with medicinal properties by using machine learning and big data analytics. (Baskin 2020)

1.7.1. Target Identification

AI systems can rapidly analyze large databases of chemical and biological data to identify potential targets for drug intervention. By drawing insights from massive amounts of readily available data, researchers are better equipped to find potential therapeutic targets that were previously undetected (Zhang, et al., 2017; Lavecchia, 2019)

1.7.2. Molecular Simulations

Molecular simulations help researchers understand how drug candidates interact with their molecular targets. AI models can predict the behavior of molecules in various scenarios, significantly reducing the time and cost of experimentation. This means that researchers can screen a large number of molecules in a matter of days and identify the most promising ones for further investigation and development (Lo et al., 2018).

1.7.3. Prediction of Drug Properties

AI models can simulate the interactions that drug candidates will have with the human body based on chemical and biological data. By analyzing this data, AI can predict drug properties such as bioavailability, toxicity, and efficacy, providing crucial information about the therapeutic potential of drug candidates. This knowledge enables scientists to make more informed decisions about which drug candidates to pursue and focus their efforts on areas with the highest likelihood of success (Gaulton, et al., 2016; Kim, et al, 2018).

1.7.4. Candidate Drug Prioritization

Once researchers have identified viable candidates, AI algorithms can rank potential therapeutic candidates by predicting their properties and comparing them with existing approved drugs. This can help researchers prioritize the most promising compounds, reducing the time and cost of bringing drugs to market (Goodfellow, 2014).

1.7.5. Synthesis Pathway Generation

AI algorithms can predict the most efficient way to synthesize therapeutic candidates. By considering the properties of drugcandidates and available resources, AI can generate synthetic routes that minimize the number of steps and reagents required whileoptimizing yields. This results in a more efficient and environmentally friendly drug production process (Kingma, 2013)

1.7.6. Drug Repurposing

Drug repurposing involves using an already-approved drug for a new therapeutic purpose. AI models can identify the potential for therapeutic repurposing by analyzing existing drug data and identifying putative off-targets. Repurposing has significant advantages because existing drugs have known safety profiles, and the development of new drugs takes less time (Hochreiter & Schmidhuber, 1997)

1.7.7. Drug Adherence and Dosage

AI algorithms that incorporate patients' genetic, physiological, and lifestyle data can help personalize drug adherence and dosage.AI can monitor a patient's response to treatment and adjust the dosage as needed to ensure the best possible therapeutic outcomes (Jumper et al., 2021).

BIOPRESS

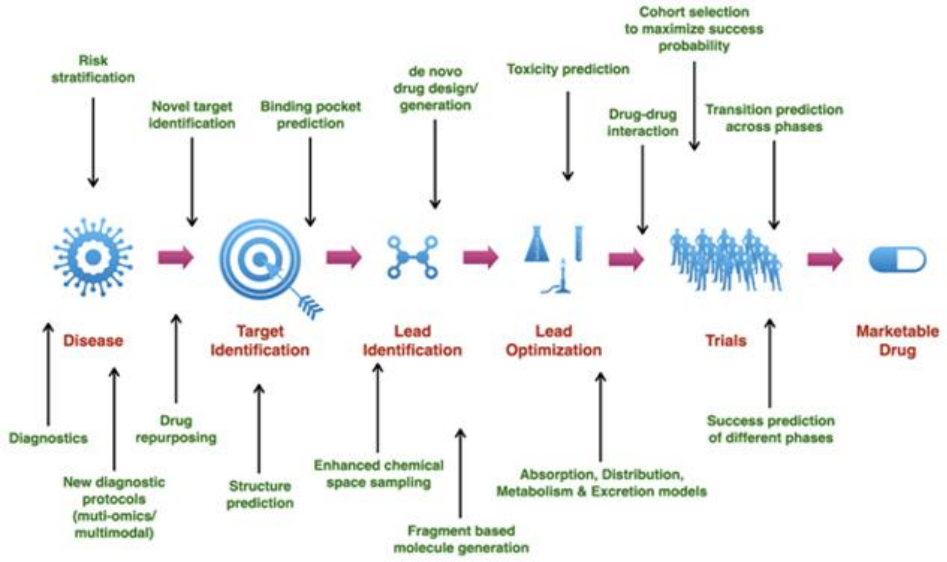


Figure 1.5: AI role in various phases of drug discovery

1.8 AI Application in the Pharmaceutical Industry

AI in Discovery of Drug Drug development and research is a costly and laborious endeavor that necessitates balancing numerous features of molecules to achieve desirable clinical efficacy and safety. Among these characteristics, the drugs' performance primarily depends on their pharmacokinetics /pharmacodynamics (PK/PD) characteristics, nonclinical safety profiles, and absorption, distribution, metabolism, and elimination (ADME) characteristics. The discovery of drug and the process of its development may benefit from recent advancements in AI/ML techniques since they could increase the likelihood of success and speed up decision making. Huge datasets that contain information about chemical structures and the associated experimental readouts are used to train the AI/ML system, which subsequently clarifies the underlying ideas, biological reactions, and interactions between chemical structures.

BIOPRESS

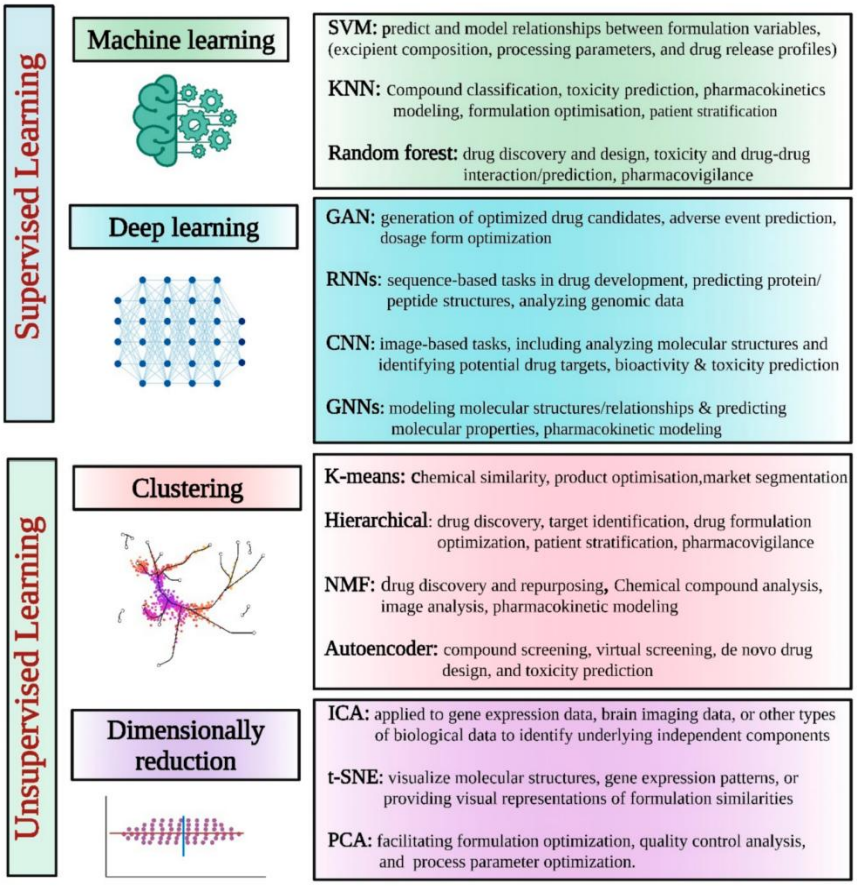

Supervised Learning
Machine learning
SVM: predict and model relationships between formulation variables, (excipient composition, processing parameters, and drug release profiles)
KNN: Compound classification, toxicity prediction, pharmacokinetics modeling, formulation optimisation, patient stratification
Random forest: drug discovery and design, toxicity and drug-drug interaction/prediction, pharmacovigilance
Deep learning
GAN: generation of optimized drug candidates, adverse event prediction, dosage form optimization
RNNs: sequence-based tasks in drug development, predicting protein/ peptide structures, analyzing genomic data
CNN: image-based tasks, including analyzing molecular structures and identifying potential drug targets, bioactivity & toxicity prediction
GNNs: modeling molecular structures/relationships & predicting molecular properties, pharmacokinetic modeling
Unsupervised Learning
Clustering
K-means: Chemical similarity, product optimisation,market segmentation
Hierarchical: drug discovery, target identification, drug formulation optimization, patient stratification, pharmacovigilance
NMF: drug discovery and repurposing, Chemical compound analysis, image analysis, pharmacokinetic modeling
Autoencoder: compound screening, virtual screening, de novo drug design, and toxicity prediction
Dimensionally reduction
ICA: applied to gene expression data, brain imaging data, or other types of biological data to identify underlying independent components
t-SNE: visualize molecular structures, gene expression patterns, or providing visual representations of formulation similarities
PCA: facilitating formulation optimization, quality control analysis, and process parameter optimization.

Figure 1.6: AI Learning Models/Tools for Pharmaceutical applications

Several uses of AI in Pharmaceutical Industry:

a Target validation and choice Discovering the role of putative drug targets (genes/proteins of a small molecule) and their importance in illness requires understanding how these molecules work. The selection and validation methods must take into account functioning genomics, structure genomics, proteomics, in-vitro cell-based assays, and in-vivo animal research assays. AI evaluates the Drug Information Bank to ascertain the beneficial effects of drug candidates, genetic expressions, interactions between proteins, and clinical data records from a public library. To ascertain a drug's efficacy target, it is required to understand the function of potential molecular targets and their significance in illness.

b. Screening of compounds and lead optimization To select drug candidates, compound screening and lead optimisation procedures combine virtual screening, high throughput screening, and combinatorial chemistry. To build the chemical database for AI-based virtual screening, significant amounts of compounds are retrieved from openly accessible chemogenomics libraries, which include millions of compounds with structural information.

c. Preclinical research Preclinical studies, sometimes referred to as non-clinical studies, are laboratory tests conducted in vitro and in vivo on novel medicinal substances to ascertain their safety and efficacy profile. To quickly collect relevant huge amounts of biological data, clustering-based ML methods assess RNA sequencing technologies to discover a molecular mechanism of action. It also exposes several previously unknown relationships between different stimuli and the cytokines they affect.

d. Clinical Research For recognizing patient conditions, finding gene targets, forecasting the outcome of a created drug, and on and off targets, a generation of AI tools for clinical trials would prove excellent. In Phase II clinical studies, one AI mobile application boosted medicine adherence by 25% compared to conventional direct observation therapy.

e. AI IN DRUG SCREENING The use of AI in drug screening includes target identification; molecular simulations; drug property predictions; de novo drug creation, synthesis of route generation, and candidate drug prioritization (Mak & Pichika, 2018).

f. AI FOR PHARMACEUTICAL PRODUCTION

Manufacturing complexity has increased from prior years as a result of recent improvements in pharmaceutical technology and distinctive developments. For the complex procedure, a proven product's excellence and effectiveness are also necessary. Modern pharmaceutical production systems are striving to merge human expertise into machines to enhance manufacturing processes. The employment of AI in the production process has the potential to expand the pharmaceutical industry. Numerous pharmaceutical activities will be mechanized with the aid of AI tools. In addition to addressing problems that may occur during the manufacturing process, this can aid in numerical stimulation17. AI can be included in the manufacturing of medications to achieve the right level of outcomes by utilizing other cutting-edge manufacturing technologies. According to a discussion paper released by the U.S. FDA, the following potential applications of AI were seen in pharmaceutical production processes (Mak & Pichika, 2018).

Uses of AI in production of new pharmaceutical products:

• In Identification & diagnosis of Disease

• Digital therapeutics & Personalized treatment and behavioral modification

• Drug Development and Production

• Prognostic forecasting

• Medical Tests

1.9 Limitations of AI in Drug Discovery

Currently, medicinal chemistry methods rely heavily on a hit-and-miss approach and large-scale testing techniques. These techniques involve examining large numbers of potential drug

compounds, in order to identify those with the desired properties. However, these methods can be slow, costly, and often yield results with low accuracy. In addition, they can be limited by the availability of suitable test compounds and the difficulty of accurately predicting their behavior in the body (Pu et al., 2019)

Furthermore, AI-based algorithms can also be employed to identify new targets for drug development, such as the specific proteins or genetic pathways involved in diseases . This can expand the scope of drug discovery beyond the limitations of more conventional approaches and may eventually lead to the development of novel and more effective medications Thus, while traditional methods of pharmaceutical research have been relatively successful in the past, they are limited by their reliance on trial-and-error experimentation and their inability to accurately predict the behavior of new potential bioactive compounds. AI-based approaches, on the other hand, have the ability to improve the efficiency and accuracy of drug discovery processes and can lead to the development of more effective medications (Liu et al, 2021; Bannigan et al, 2021).

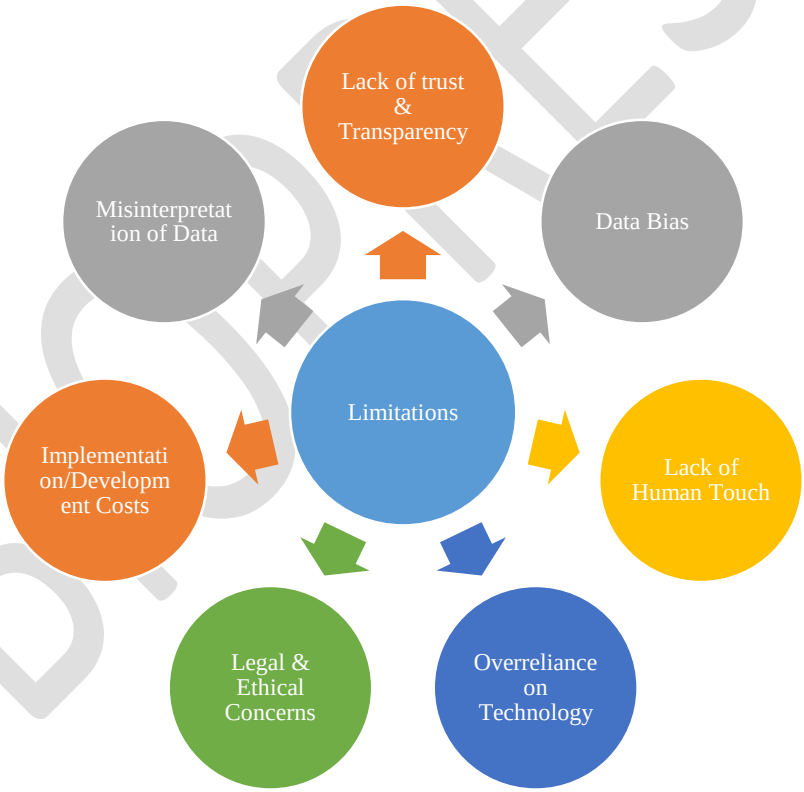


Figure 1.7: Limitations of AI in Drug Design

1.10 AI in Clinical Drug Trials

Artificial intelligence (AI), a branch of computer science, seeks to understand how the human brain tackles problems and reaches decisions . It has existed for nearly a century, and its novelty doesn't stem from its applications. In recent decades, drug discovery has grown increasingly intricate, with AI playing a supporting role—yet this contribution rarely gets much attention.

Artificial intelligence (AI) offers a chance to rethink the clinical trial ecosystem. By harnessing advances in machine learning (ML), natural language processing (NLP), and generative AI (such as large language models [LLMs]), investigators can create more adaptive protocols, identify and recruit eligible participants more effectively, and streamline trial monitoring along with endpoint assessment. Beyond mere automation, AI enables the integration of vast data volumes and incorporates multimodal capabilities.

Clinical trials stand as one of the most vital yet demanding phases in drug development and medical advancement. Historically, they've grappled with issues like extended durations, steep expenses, hurdles in enrolling patients, flawed protocols, and intricate regulations. With healthcare shifting toward precision therapies and data-centric studies, AI and machine learning are revolutionizing the design, conduct, and refinement of clinical trials.

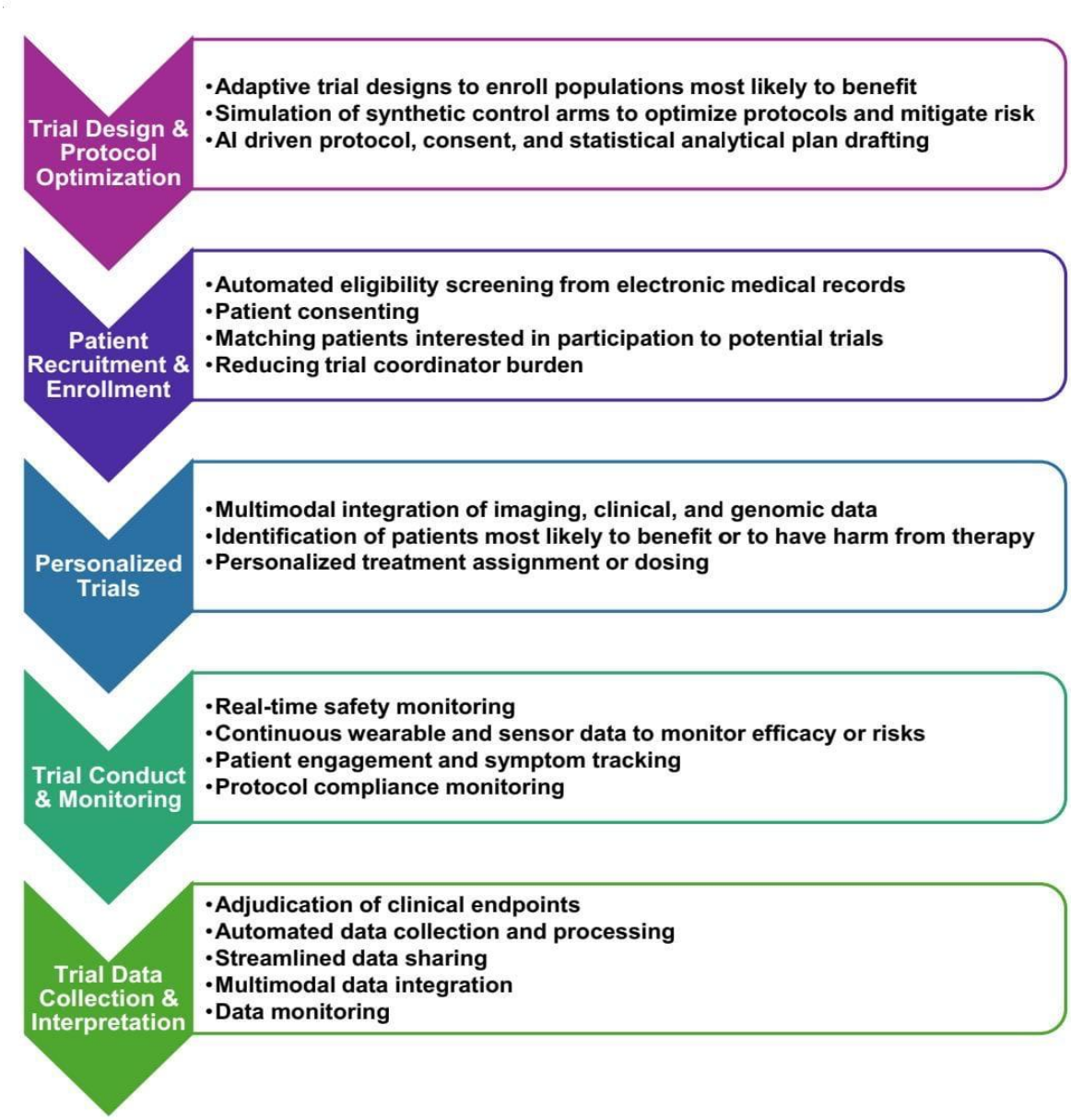

Trial Design & Protocol Optimization
•Adaptive trial designs to enroll populations most likely to benefit
•Simulation of synthetic control arms to optimize protocols and mitigate risk
•AI driven protocol, consent, and statistical analytical plan drafting
Patient Recruitment & Enrollment
•Automated eligibility screening from electronic medical records
•Patient consenting
•Matching patients interested in participation to potential trials
•Reducing trial coordinator burden
Personalized Trials
•Multimodal integration of imaging, clinical, and genomic data
•Identification of patients most likely to benefit or to have harm from therapy
•Personalized treatment assignment or dosing
Trial Conduct & Monitoring
•Real-time safety monitoring
•Continuous wearable and sensor data to monitor efficacy or risks
•Patient engagement and symptom tracking
•Protocol compliance monitoring
Trial Data Collection & Interpretation
•Adjudication of clinical endpoints
•Automated data collection and processing
•Streamlined data sharing
•Multimodal data integration
•Data monitoring

Figure 1.8: Use of AI in clinical trials

1.11 Applications of AI to clinical trials:-

1 Optimizing Trial Design & Protocols:-

Designing a clinical trial protocol is a pivotal step in the trial process, ensuring not only the robustness of the scientific inquiry but also the feasibility, safety, and broad applicability of the outcomes. Traditional protocol development leans heavily on expert consensus and limited data, which can result in inefficiencies, redundancies, and avoidable trial failures.

2 Biosimulation:-

Biosimulation involves using computers to model and replicate biological systems and processes through mathematical frameworks. By integrating AI and machine learning with models and data, these simulations produce forecasts about a drug's impacts, helping researchers gain deeper insights.

3 Recruiting patients through data mining:-

Patient recruitment through data mining plays a crucial role in clinical trials and drug development. Trials often fail when research organizations struggle to enroll sufficient participants promptly or when recruitment strategies prove ineffective. AI can analyze population data to identify ideal candidates who stand to benefit most from participating in clinical studies. Additionally, by examining data to locate regions or countries with the highest disease prevalence. AI tools can optimize the selection of recruitment sites.

4 Methods of early diseases diagnosis:-

Medical facilities and research centers are increasingly creating algorithms for early disease detection. These programs analyze symptoms, patient histories, and testing procedures. The resulting algorithm can then assess if a patient is at risk of or in the early stages of a disease.

1.12 References

1. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*,*1–2*, 19–25 https://doi.org/10.1016/j.softx.2015.06.001

2. Alpaydin, E. (2021). *Machine learning*. MIT press.

3. Bannigan, P., Aldeghi, M., Bao, Z., Häse, F., Aspuru-Guzik, A., & Allen, C. (2021). Machine learning directed drug formulation development. *Advanced Drug Delivery Reviews*,*175*, 113806.https://doi.org/10.1016/j.addr.2021.05.016

4. Baskin, I. I. (2020). The power of deep learning to ligand-based novel drug discovery. *Expert Opinion on Drug Discovery*,*15*(7),755–764. https://doi.org/10.1080/17460441.2020.1745183

5. Buterez, D., Janet, J. P., Kiddle, S. J., & Liò, P. (2023). MF-PCBA: Multifidelity High-Throughput Screening Benchmarks for Drug Discovery and Machine Learning.*Journal of Chemical Information and Modeling*, *63*(9), 2667–2678. https://doi.org/10.1021/acs.jcim.2c01569

6. Buterez, D., Janet, J. P., Kiddle, S. J., Oglic, D., & Lió, P. (2024). Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting. *Nature Communications*, *15*(1), 1517. https://doi.org/10.1038/s41467-024-45566-8

7. Fetzer, J. H. (1990). What is Artificial Intelligence? In *Studies in cognitive systems* (pp. 3–27). https://doi.org/10.1007/978-94-009-1900-6_1

8. Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M. P., Overington, J. P., Papadatos, G., Smit, I., & Leach, A. R. (2016). The ChEMBL database in 2017. *Nucleic Acids Research*, *45*(D1), D945–D954. https://doi.org/10.1093/nar/gkw1074

9. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2, pp. 1-800). Cambridge: MIT press.

10. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative adversarial Nets*. https://proceedings.neurips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

11. Gurney, K. (2018). *An introduction to neural networks*. CRC press.

12. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

13. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

14. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. a. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

15. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*, *3*(11), 935–949. https://doi.org/10.1038/nrd1549

16. Kingma, D. P., & Welling, M. (2013, December 20). *Auto-Encoding variational Bayes*. arXiv.org. https://arxiv.org/abs/1312.6114

17. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2018). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*, *47*(D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033

18. Lavecchia, A. (2019). Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discovery Today*, *24*(10), 2017–2032. https://doi.org/10.1016/j.drudis.2019.07.006

19. Lim, J., Ryu, S., Kim, J. W., & Kim, W. Y. (2018). Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of Cheminformatics*, *10*(1), 31. https://doi.org/10.1186/s13321-018-0286-7

BIOPRESS

20. Lo, Y., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today*, *23*(8), 1538–1546. https://doi.org/10.1016/j.drudis.2018.05.010

21. Liu, X., IJzerman, A. P., & Van Westen, G. J. P. (2020). Computational Approaches for de novo drug Design: past, present, and future. *Methods in Molecular Biology*, *2190*, 139–165.https://doi.org/10.1007/978-1-0716-0826-5_6

22. Meyers, J., Fabian, B., & Brown, N. (2021). De novo molecular design and generative models. *Drug Discovery Today*,*26*(11), 2707–2715.https://doi.org/10.1016/j.drudis.2021.05.019

23. Mak, K. K., Wong, Y. H., & Pichika, M. R. (2024). Artificial intelligence in drug discovery and development. *Drug discovery and evaluation: safety and pharmacokinetic assays*, 1461-1498.

24. Miles, J., & Walker, A. (2006). The potential application of artificial intelligence in transport.*IEE Proceedings Intelligent Transport Systems*,*153*(3), 183. https://doi.org/10.1049/ip-its:20060014

25. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791.https://doi.org/10.1002/jcc.21256

26. Mak, K., & Pichika, M. R. (2018). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, *24*(3), 773–780. https://doi.org/10.1016/j.drudis.2018.11.014

27. Pu, L., Naderi, M., Liu, T., Wu, H., Mukhopadhyay, S., & Brylinski, M. (2019). eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates.*BMC Pharmacology and Toxicology*,*20*(1).https://doi.org/10.1186/s40360-018-0282-6

28. Ramesh, A., Kambhampati, C., Monson, J., & Drew, P. (2004). Artificial intelligence in medicine. *Annals of the Royal College of Surgeons of England*, *86*(5), 334–338. https://doi.org/10.1308/147870804290

29. Sarella, P. N. K., Vipparthi, A. K., Valluri, S., Vegi, S., & Vendi, V. K. (2024). Nanorobotics: pioneering drug delivery and development in pharmaceuticals.*Research Journal of Pharmaceutical Dosage Forms and Technology*, 81–90.https://doi.org/10.52711/0975-4377.2024.00014

30. Tropsha, A. (2010). Best Practices for QSAR model development, Validation, and Exploitation.*Molecular Informatics*,*29*(6–7), 476–488.https://doi.org/10.1002/minf.201000061

31. Vora, L. K., Gholap, A. D., Jetha, K., Thakur, R. R. S., Solanki, H. K., & Chavda, V. P. (2023). Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics*,*15*(7), 1916.https://doi.org/10.3390/pharmaceutics15071916

32. Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial Intelligence and the Public Sector—Applications and Challenges.*International Journal of Public Administration*,*42*(7), 596–615.https://doi.org/10.1080/01900692.2018.1498103

33. Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Abel, R. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern Free-Energy calculation protocol and force field. *Journal of the American Chemical Society*,*137*(7), 2695–2703.https://doi.org/10.1021/ja512751q

34. Wei, J., Zhuo, L., Fu, X., Zeng, X., Wang, L., Zou, Q., & Cao, D. (2024). DrugReAlign: a multisource prompt framework for drug repurposing based on large language models. *BMC Biology*,*22*(1), 226.https://doi.org/10.1186/s12915-024-02028-3

35. Yang, Y., & Siau, K. L. ().*A qualitative research on marketing and sales in the Artificial intelligence age*. AIS Electronic Library (AISeL).https://aisel.aisnet.org/mwais2018/41/

36. Ye, Q., Hsieh, C., Yang, Z., Kang, Y., Chen, J., Cao, D., He, S., & Hou, T. (2021). A unified drug–target interaction prediction framework based on knowledge graph and recommendation system.*Nature Communications*,*12*(1), 6775.https://doi.org/10.1038/s41467-021-27137-3

37. Zhou, S. F., & Zhong, W. Z. (2017). Drug design and discovery: principles and applications. *Molecules*, *22*(2), 279.

38. Zende, A. K. D. M. S. A. (2025). Drug Design: A Comprehensive review. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.14774296

39. Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery.*Drug Discovery Today*, *22*(11), 1680–1685.https://doi.org/10.1016/j.drudis.2017.08.010

BIOPRESS

Chapter 2

Structure-Based Drug Design and Molecular Docking Approaches in Contemporary Therapeutic Innovation

Jagruti Gosavi*, Rajeshwari Dughde, Trupti Kadam

K. V. N. Naik S. P. Sanstha's, Institute of Pharmaceutical Education & Research, Canada Corner, Nashik, 422002, Maharashtra, India.

Abstract

Structure-Based Drug Design (SBDD) has fundamentally transformed the paradigm of modern drug discovery by enabling rational, target-driven development of therapeutic agents. The approach leverages high-resolution structural information of biological macromolecules to design small molecules with enhanced specificity, affinity, and pharmacological efficacy. Advances in structural biology techniques—including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM)—have provided unprecedented insight into protein architecture and ligand-binding mechanisms. Complementing these experimental methods, molecular docking has emerged as a central computational strategy for predicting ligand binding modes and estimating binding affinities.

This chapter presents an in-depth exploration of the principles, methodologies, computational frameworks, and translational applications of SBDD and molecular docking. Topics include thermodynamic foundations of molecular recognition, structural biology techniques, target identification and validation, binding site analysis, docking algorithms, scoring functions, protein flexibility modeling, molecular dynamics simulations, free energy calculations, virtual screening workflows, fragment-based drug design, covalent docking strategies, and artificial intelligence-assisted drug discovery. Case studies of clinically approved drugs developed using structure-based approaches are discussed to illustrate real-world impact. The chapter concludes by addressing limitations, challenges, and future perspectives that integrate structural biology, computational chemistry, and machine learning to accelerate precision medicine

Keywords

Structure-Based Drug Design; Molecular Docking; Protein–Ligand Interaction; Virtual Screening; Binding Free Energy; Molecular Dynamics; Cryo-EM; Artificial Intelligence in Drug Discovery; Fragment-Based Design; Computational Pharmacology

BIOPRESS

2.1. Introduction

Drug discovery has evolved significantly over the past century. Early pharmaceutical development relied heavily on natural product isolation and empirical screening approaches. Although these strategies yielded important therapeutic breakthroughs, they often lacked mechanistic insight into drug action at the molecular level.

The integration of molecular biology, structural biochemistry, and computational modeling marked a paradigm shift toward rational drug discovery. Structure-Based Drug Design (SBDD) emerged as a powerful methodology that uses three-dimensional (3D) structural information of biological targets to guide the design and optimization of small-molecule therapeutics.

Unlike traditional screening methods that test thousands of compounds experimentally, SBDD allows researchers to predict how molecules interact with specific target sites before synthesis. Molecular docking, a core component of SBDD, computationally simulates ligand–receptr interactions and ranks compounds based on predicted binding affinity.

With the availability of structural databases such as the Protein Data Bank (PDB) and advanced computational resources, SBDD has become indispensable in pharmaceutical research.

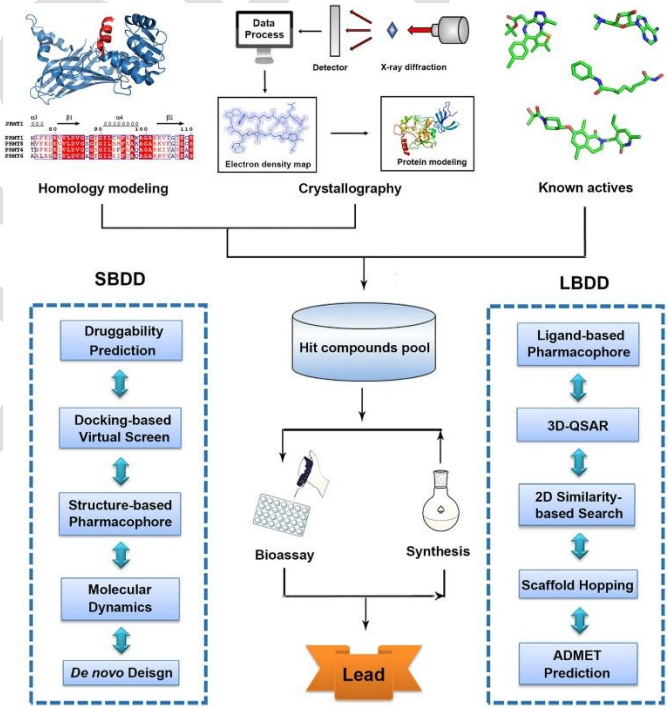


Figure 2.1: interactions and ranks compounds based on predicted binding affinity.

2.2. Historical Development of Structure-Based Drug Design

The conceptual foundation of SBDD dates back to Emil Fischer's "lock-and-key" hypothesis in 1894, which proposed that molecular complementarity governs enzyme–substrate interactions. In the mid-20th century, advances in enzymology and protein chemistry expanded understanding of molecular recognition.

Key milestones include:

1958: First protein structure (myoglobin) determined by X-ray crystallography

1980s: Introduction of computer-aided drug design tools

1990s: Development of automated molecular docking programs

2000s: Growth of high-throughput virtual screening

2010s: Cryo-EM revolution and increased structural resolution

2020s: Artificial intelligence-based protein structure prediction (e.g., AlphaFold)

The approval of drugs such as Captopril (ACE inhibitor) and HIV protease inhibitors demonstrated the clinical success of structure-based strategies.

2.3. Thermodynamic and Molecular Basis of Drug–Target Interaction

2.3.1 Binding Free Energy

Ligand binding is governed by thermodynamic principles. The Gibbs free energy equation describes the spontaneity of binding:

\Delta G_{bind} = \Delta H - T\Delta S

Where:

ΔH: Enthalpic contribution (interaction energy)

ΔS: Entropic contribution (disorder changes)

T: Temperature

A negative ΔG indicates favorable binding.

BIOPRESS

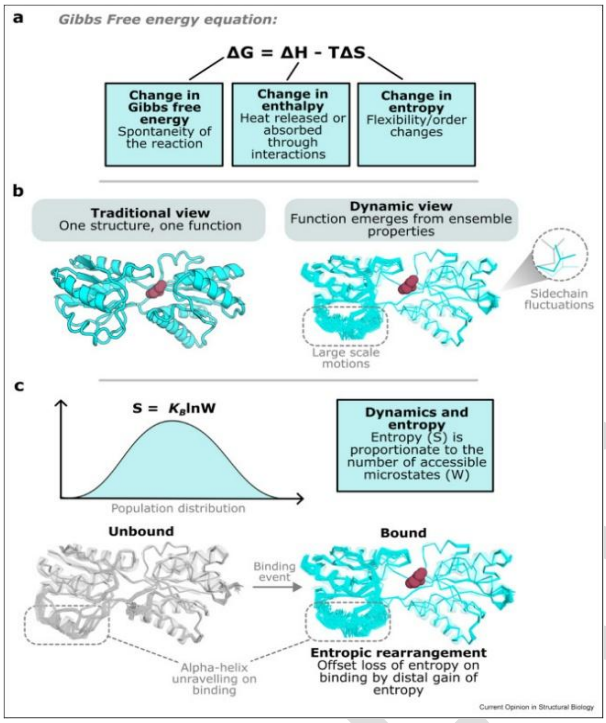


Figure 2.2: gibbs free energy Equation

2.3.2 Intermolecular Forces

Drug–target interactions involve:
Hydrogen bonding
Electrostatic interactions
Hydrophobic effects
Van der Waals forces
π–π interactions
Halogen bonding
These interactions collectively stabilize the ligand–receptor complex.

2.3.3 Solvent Effects and Desolvation

Water molecules influence binding through:
Solvent displacement
Water-mediated hydrogen bonds
Hydrophobic collapse
Modeling solvent effects accurately remains computationally challenging.

2.4. Target Identification and Validation

2.4.1 Target Identification

Drug targets are typically proteins associated with disease pathology:

Enzymes
Receptors
Ion channels
Transport proteins
Viral proteins
Genomic and proteomic technologies aid in identifying novel targets.

2.4.2 Target Validation

Validation ensures therapeutic relevance. Techniques include:
Gene knockout models
RNA interference
CRISPR-Cas9 editing
Biomarker studies
Pharmacological inhibition assays

2.5. Structural Biology Techniques

2.5.1 X-Ray Crystallography

Provides high-resolution atomic structures. Advantages include precise coordinate determination. Limitations include crystallization difficulty and static representation.

2.5.2 Nuclear Magnetic Resonance (NMR)

Captures dynamic behavior and solution-phase structures. Suitable for small proteins.

5.3 Cryo-Electron Microscopy

Enables structural determination of large complexes and membrane proteins without crystallization.

2.5.4 Computational Modeling

Includes:

Homology modeling
Threading
Ab initio prediction
AlphaFold-based AI modelling

BIOPRESS

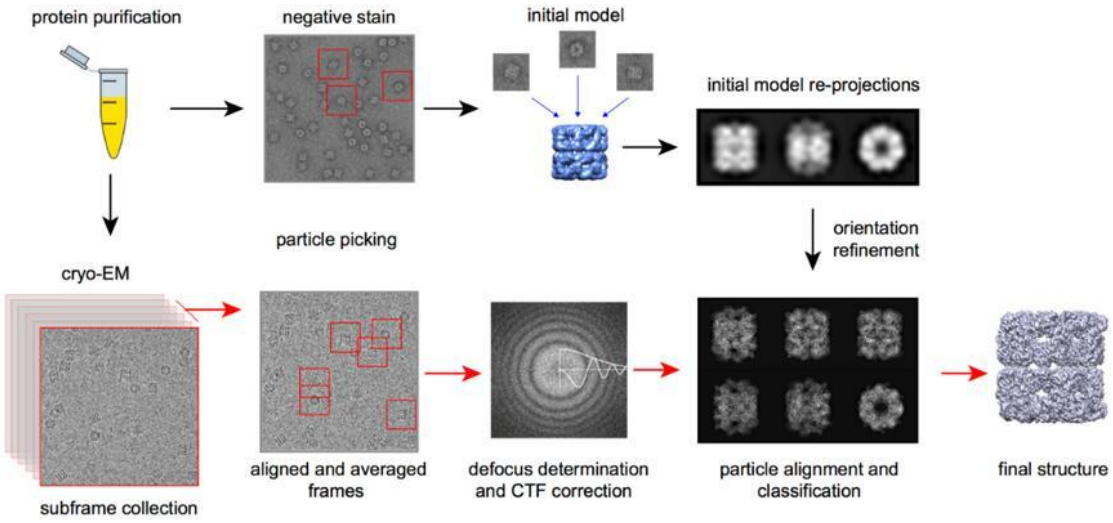


Figure 2.3: Methods for Determining Protein Structures in Drug Discovery

2.6. Binding Site Identification and Characterization

Active site identification involves:
Analysis of co-crystallized ligands
Pocket detection software
Site-directed mutagenesis
Surface mapping
Key properties assessed:
Volume and depth
Hydrophobicity
Electrostatic potential
Residue conservation

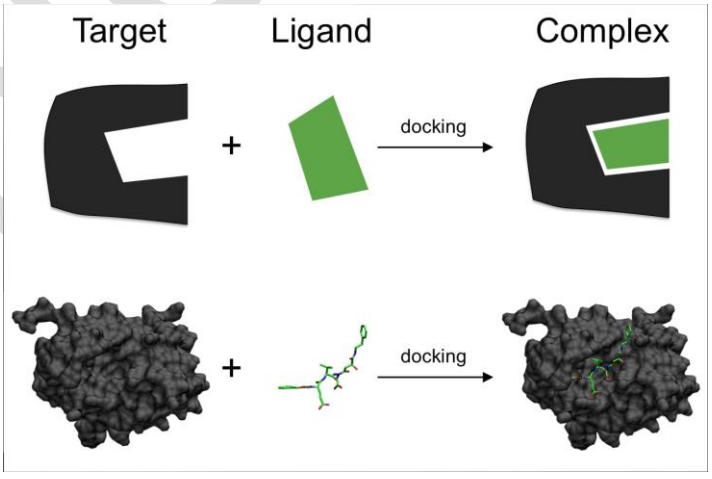


Figure 2.4: Representation of a Protein Binding Pocket with Key Interactions

2.7. Molecular Docking: Algorithms and Methodologies

Docking predicts ligand binding orientation and affinity.

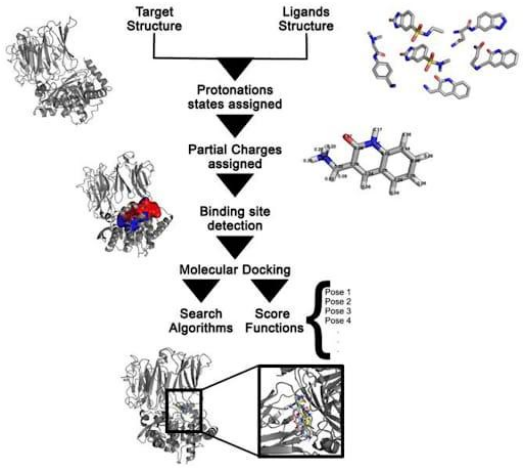


Figure 2.5: Molecular Docking Workflow

2.7.1 Search Algorithms

Efficient sampling is critical.
Systematic Search
Exhaustive but computationally demanding.
Genetic Algorithms
Mimic evolutionary processes.
Monte Carlo Methods
Random sampling with probabilistic acceptance.
Simulated Annealing
Gradual energy minimization strategy

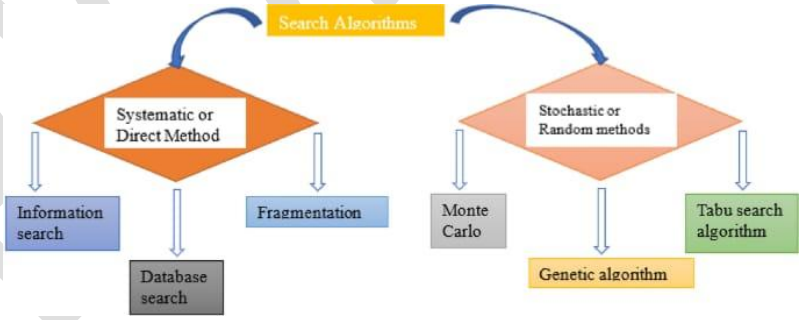


Figure 2.6: Search Algorithm Strategies in Molecular Docking
Qiagram showing:

2.7.2 Scoring Functions

Force-Field Based
Based on classical energy terms.
Empirical
Weighted interaction parameters.

Knowledge-Based
Statistical potentials from structural data.
Machine Learning-Based

AI models trained on binding datasets.

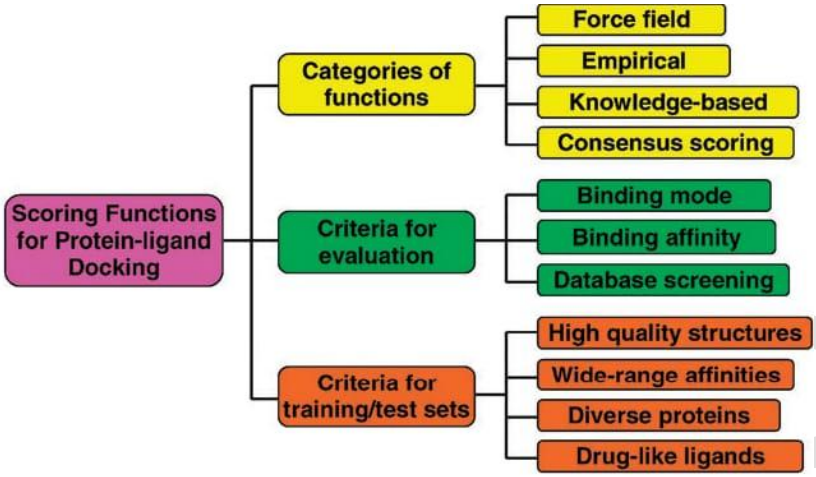


Figure 2.7: Types of Scoring Functions in Docking

2.8. Protein Flexibility and Advanced Docking

Rigid docking may neglect protein conformational changes.

Advanced methods:
Flexible docking
Induced-fit docking
Ensemble docking
Covalent docking
Fragment-based docking

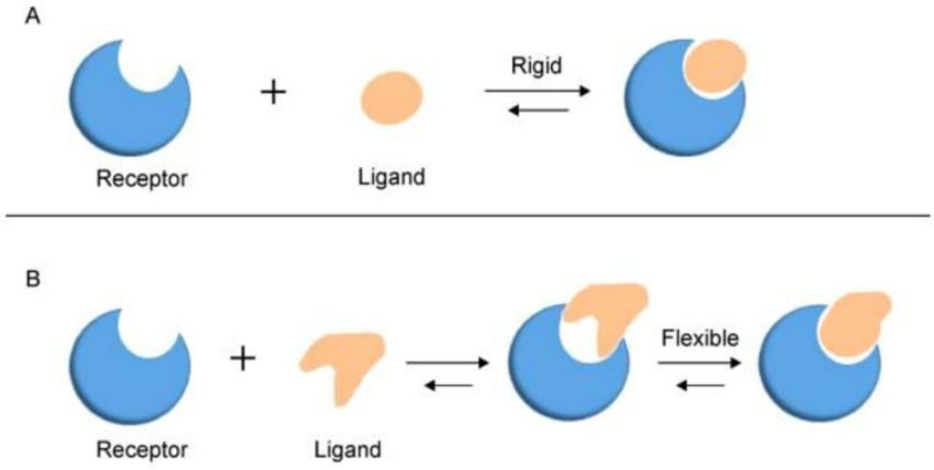


Figure 2.8: Rigid Docking vs Flexible Docking

2.9. Molecular Dynamics Simulations

MD simulations incorporate time-dependent behavior.
Applications:

Stability assessment
Conformational exploration
Solvent interaction analysis
RMSD and RMSF calculations

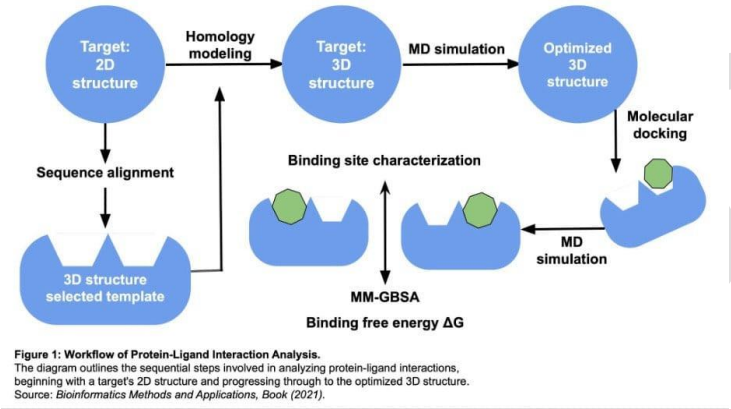

Target: 2D structure
Homology modeling
Target: 3D structure
MD simulation
Optimized 3D structure
Molecular docking
Sequence alignment
Binding site characterization
3D structure selected template
MD simulation
MM-GBSA
Binding free energy ΔG
Figure 1: Workflow of Protein-Ligand Interaction Analysis.
The diagram outlines the sequential steps involved in analyzing protein-ligand interactions, beginning with a target's 2D structure and progressing through to the optimized 3D structure.
Source: Bioinformatics Methods and Applications, Book (2021).

Figure 2.9: Molecular Dynamics Simulation Cycle

2.10. Binding Free Energy Calculations

Accurate energy estimation methods include:

MM-PBSA

MM-GBSA

Binding Free Energy Calculations

Accurate energy estimation methods include:

MM-PBSA

MM-GBSA

Free Energy Perturbation (FEP)

Thermodynamic Integration (TI)

These improve reliability beyond docking scores.

Free Energy Perturbation (FEP)

Thermodynamic Integration (TI)

These improve reliability beyond docking scores.

2.11. Virtual Screening and Lead Optimization

Virtual screening workflow:

1. Target preparation
2. Ligand library preparation
3. Docking
4. Scoring and ranking
5. ADMET prediction
6. Experimental validation

Lead optimization involves iterative structural modification guided by docking and simulation.

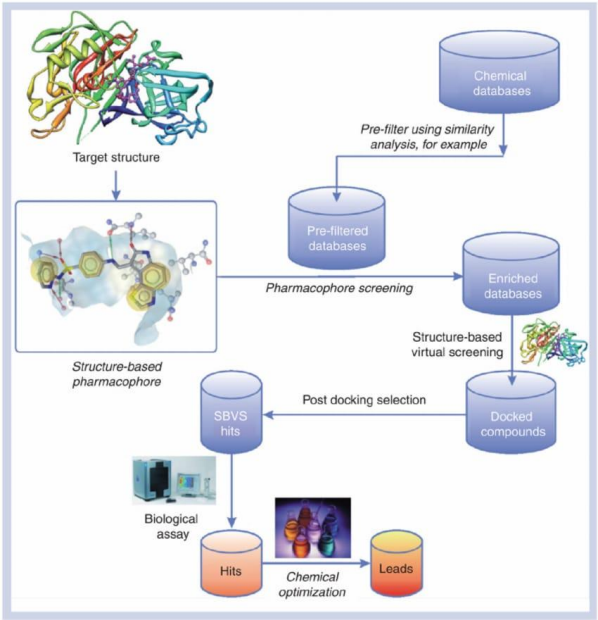


Figure 2.10: Structure-Based Virtual Screening Pipeline

2.12. Pplications of Structure-Based Drug Design

Oncology

Kinase inhibitors (Imatinib, Vemurafenib)

Infectious Diseases

HIV protease inhibitors, SARS-CoV-2 Mpro inhibitors

Cardiovascular Diseases

ACE inhibitors, renin inhibitors

Neurological Disorders

GPCR modulators

Antimicrobial Resistance

β-lactamase inhibitors

2.13. Case Studies

Captopril – First rationally designed ACE inhibitor

Saquinavir – HIV protease inhibitor

Imatinib – BCR-ABL kinase inhibitor

Oseltamivir – Influenza neuraminidase inhibitor

Vemurafenib – BRAF inhibitor

2.14. Challenges and Limitations

Protein flexibility modeling limitations

Solvent and entropy approximation errors

False positives in docking

Resistance mutations

Computational resource requirements

2.15. Emerging Trends and Future Directions

Artificial intelligence-enhanced docking

Deep learning scoring functions

Cryo-EM-guided drug design

Quantum mechanics/molecular mechanics (QM/MM)

Personalized structure-based therapeutics

Cloud-based distributed screening

2.16. Conclusion

Structure-Based Drug Design and molecular docking have revolutionized drug discovery by enabling precise, rational targeting of disease-associated proteins. Through integration of structural biology, computational chemistry, molecular dynamics, and artificial intelligence, SBDD provides a powerful platform for accelerating therapeutic development. Although challenges remain in modeling protein flexibility, solvent dynamics, and entropic contributions, ongoing advancements in computational algorithms and structural determination methods continue to improve predictive accuracy. As structural datasets expand and machine learning techniques mature, SBDD will play an increasingly central role in the development of personalized and precision medicines.

2.17 References

1. Anderson, A. C. (2003). The process of structure-based drug design. Chemistry & Biology, 10(9), 787–797.

BIOPRESS

2. Berman, H. M., Westbrook, J., Feng, Z., et al. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242.

3. Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. Journal of Medicinal Chemistry, 43(25), 4759–4767.*

4. Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery. Current Topics in Medicinal Chemistry, 9(16), 1524–1535.*

5. Congreve, M., Chessari, G., Tisi, D., & Woodhead, A. J. (2008). Recent developments in fragment-based drug discovery. Journal of Medicinal Chemistry, 51(13), 3661–3680.*

6. Ewing, T. J. A., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking. Journal of Computer-Aided Molecular Design, 15(5), 411–428.*

7. Ferreira, L. G., dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. Molecules, 20(7), 13384–13421.*

8. Forli, S., Huey, R., Pique, M. E., et al. (2016). Computational protein–ligand docking and virtual drug screening with AutoDock Vina. Nature Protocols, 11(5), 905–919.*

9. Friesner, R. A., Banks, J. L., Murphy, R. B., et al. (2004). Glide: A new approach for rapid, accurate docking. Journal of Medicinal Chemistry, 47(7), 1739–1749.*

10. Goodsell, D. S., & Olson, A. J. (1990). Automated docking of substrates to proteins. Proteins, 8(3), 195–202.*

11. Gohlke, H., & Case, D. A. (2004). Converging free energy estimates: MM-PBSA and MM-GBSA. Journal of Computational Chemistry, 25(2), 238–250.*

12. Halgren, T. A., Murphy, R. B., Friesner, R. A., et al. (2004). Glide: Docking scoring function. Journal of Medicinal Chemistry, 47(7), 1750–1759.*

13. Harder, E., Damm, W., Maple, J., et al. (2016). OPLS3 force field for drug-like molecules. Journal of Chemical Theory and Computation, 12(1), 281–296.*

14. Huang, S. Y., & Zou, X. (2010). Advances and challenges in protein–ligand docking. International Journal of Molecular Sciences, 11(8), 3016–3034.*

15. Irwin, J. J., & Shoichet, B. K. (2005). ZINC database for virtual screening. Journal of Chemical Information and Modeling, 45(1), 177–182.*

16. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589.*

17. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening. Nature Reviews Drug Discovery, 3(11), 935–949.*

18. Kollman, P. (1993). Free energy calculations: Applications to drug design. Chemical Reviews, 93(7), 2395–2417.*

BIOPRESS

19. Kuntz, I. D., Blaney, J. M., Oatley, S. J., et al. (1982). A geometric approach to macromolecule–ligand interactions. Journal of Molecular Biology, 161(2), 269–288.*

20. Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening. Current Topics in Medicinal Chemistry, 14(16), 1923–1938.*

21. Lounnas, V., Ritschel, T., Kelder, J., et al. (2013). Current progress in structure-based rational drug design. Computational and Structural Biotechnology Journal, 5(6), e201302011.*

22. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking overview. Current Computer-Aided Drug Design, 7(2), 146–157.*

23. Morris, G. M., Huey, R., Lindstrom, W., et al. (2009). AutoDock4 and AutoDockTools4. Journal of Computational Chemistry, 30(16), 2785–2791.*

24. Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking. Biophysical Reviews, 9(2), 91–102.*

25. Paulsen, J. L., et al. (2022). Cryo-EM in drug discovery. Nature Reviews Drug Discovery, 21, 1–19.*

26. Pantsar, T., & Poso, A. (2018). Binding affinity via docking: Fact and fiction. Molecules, 23(8), 1899.*

27. Schrödinger Release. (2020). Schrödinger suite 2020. Schrödinger LLC.

28. Shoichet, B. K. (2004). Virtual screening of chemical libraries. Nature, 432(7019), 862–865.*

29. Smith, R. D., et al. (2011). Ensemble docking in drug discovery. Accounts of Chemical Research, 44(10), 1307–1315.*

30. Sousa, S. F., Ribeiro, A. J. M., Coimbra, J. T. S., et al. (2013). Protein–ligand docking methods. Proteins, 81(4), 706–727.*

31. Trott, O., & Olson, A. J. (2010). AutoDock Vina improving speed and accuracy. Journal of Computational Chemistry, 31(2), 455–461.*

32. Wang, R., Fang, X., Lu, Y., & Wang, S. (2004). Comparative evaluation of scoring functions. Journal of Medicinal Chemistry, 47(12), 2977–2980.*

33. Warren, G. L., Andrews, C. W., Capelli, A. M., et al. (2006). Critical assessment of docking programs. Journal of Medicinal Chemistry, 49(20), 5912–5931.*

34. Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011). SwissParam force field parameters. Journal of Computational Chemistry, 32(11), 2359–2368.*

35. Aldeghi, M., Heifetz, A., Bodkin, M. J., et al. (2016). Absolute binding free energy calculations. Chemical Science, 7, 207–218.*

36. Baell, J. B., & Holloway, G. A. (2010). PAINS filters in drug discovery. Journal of Medicinal Chemistry, 53(7), 2719–2740.*

37. Boehr, D. D., Nussinov, R., & Wright, P. E. (2009). Dynamic conformational ensembles. Nature Chemical Biology, 5, 789–796.*

BIOPRESS

38. Brown, D. G., & Boström, J. (2016). Analysis of medicinal chemistry trends. Journal of Medicinal Chemistry, 59(10), 4443–4458.*

39. Chen, Y. C. (2015). Beware of docking pitfalls. Drug Discovery Today, 20(3), 334–343.*

40. De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of molecular dynamics in drug discovery. Journal of Medicinal Chemistry, 59(9), 4035–4061.*

41. Ferreira, R. S., et al. (2010). Fragment-based drug design review. Future Medicinal Chemistry, 2(2), 207–222.*

42. Gorgulla, C., et al. (2020). High-throughput virtual screening at scale. Nature, 580, 663–668.*

43. Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics in drug discovery. Neuron, 99(6), 1129–1143.*

44. Kitchen, D. B., & Bajorath, J. (2010). Advances in virtual screening. Future Medicinal Chemistry, 2(6), 949–964.*

45. Klebe, G. (2013). Applying thermodynamic profiling in lead finding. Nature Reviews Drug Discovery, 14, 95–110.*

46. Lyu, J., et al. (2019). Ultra-large library docking for drug discovery. Nature, 566, 224–229.*

47. McInnes, C. (2007). Virtual screening strategies in drug discovery. Current Opinion in Chemical Biology, 11(5), 494–502.*

48. Mobley, D. L., & Gilson, M. K. (2017). Predicting binding free energies. Annual Review of Biophysics, 46, 531–558.*

49. Mysinger, M. M., & Shoichet, B. K. (2010). Rapid context-dependent ligand desolvation scoring. Journal of Chemical Information and Modeling, 50(9), 1561–1573.*

50. Nussinov, R., Tsai, C. J., & Jang, H. (2019). Protein flexibility and drug discovery. Trends in Pharmacological Sciences, 40(12), 889–902.*

51. Paul, S. M., et al. (2010). Improving R&D productivity. Nature Reviews Drug Discovery, 9, 203–214.*

52. Rognan, D. (2017). Structure-based approaches in drug discovery. Journal of Chemical Information and Modeling, 57(3), 481–492.*

53. Schames, J. R., et al. (2004). Induced fit in molecular docking. Journal of Medicinal Chemistry, 47(8), 1879–1881.*

54. Sunseri, J., & Koes, D. R. (2016). Pharmit interactive virtual screening. Nucleic Acids Research, 44(W1), W442–W448.*

55. Vamathevan, J., et al. (2019). Applications of machine learning in drug discovery. Nature Reviews Drug Discovery, 18, 463–477.*

BIOPRESS

56. Walters, W. P., & Murcko, M. A. (2020). AI in medicinal chemistry. Journal of Medicinal Chemistry, 63(16), 8655–8660.*

57. Wang, L., et al. (2015). Accurate free energy calculations for drug design. Journal of the American Chemical Society, 137(7), 2695–2703.*

58. Zhavoronkov, A., et al. (2019). AI-designed drug discovery. Nature Biotechnology, 37, 1038–1040.*

BIOPRESS

Chapter 3

Lead Optimization Strategies in Small Molecule Development

Ashwini Badhe

Swalife Foundation, Pune, India

Abstract

The pharmaceutical drug discovery pipeline is an arduous and resource-intensive endeavor, traditionally characterized by high attrition rates in the clinical stages. Lead optimization represents the most critical phase of this process, acting as the bridge between the identification of initial hits and the nomination of a preclinical candidate. This report provides an exhaustive analysis of the strategies employed in modern lead optimization to refine small molecules for therapeutic use. Central to this phase is the simultaneous improvement of potency, selectivity, and drug-likeness. The analysis delves into foundational medicinal chemistry principles such as Structure-Activity Relationship (SAR) analysis, bioisosterism, and scaffold hopping, while also examining advanced computational methodologies including Structure-Based Drug Design (SBDD) and Ligand-Based Drug Design (LBDD).

A significant portion of the report is dedicated to the optimization of absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles, which remain the primary hurdles for clinical success. We examine the role of Multi-Parameter Optimization (MPO) in navigating the complex trade-offs between conflicting molecular properties. Furthermore, the transformative impact of artificial intelligence (AI) and machine learning (ML) is explored, showcasing how generative models and predictive algorithms are compressing development timelines and reducing the economic burden of drug discovery. Case studies, including the development of the inhibitor Venetoclax and the AI-driven discovery of ISM001-055, illustrate the practical application of these strategies. Finally, emerging trends such as microfluidics-based screening, green chemistry, and targeted protein degradation are discussed, providing a forward-looking perspective on the future of small molecule development. By integrating diverse experimental and computational tools, lead optimization aims to enhance the quality of

BIOPRESS

compounds entering clinical trials, ultimately improving patient outcomes through safer and more effective therapies.

Keywords

Lead optimization; Small molecules; SAR (Structure–Activity Relationship); ADMET; Medicinal chemistry; Drug design; Multi-parameter optimization; Artificial intelligence; Fragment-based drug discovery; Scaffold hopping.

3.1 Introduction

The evolution of a new therapeutic agent is a complex journey that initiates with the discovery of a biological target and culminates in the regulatory approval of a safe and effective drug. Within the drug discovery pipeline, the transition from early-stage screening to clinical readiness is governed by a series of rigorous optimization steps. The process begins with "Hits"—compounds that demonstrate a reproducible, though often weak, interaction with a target of interest in high-throughput screening (HTS) or phenotypic assays. These hits are further validated and expanded into "Leads," which possess a confirmed chemical scaffold and a preliminary understanding of their structure-activity profile. Lead optimization is the subsequent, iterative phase where these leads are meticulously refined to meet the stringent criteria of a "Preclinical Candidate" (PCC).

In small molecule development, lead optimization is not a linear path but a multi-dimensional balancing act. The medicinal chemist must improve the "Potency" of the molecule—its ability to produce a desired effect at low concentrations—while ensuring "Selectivity" to avoid off-target interactions that could lead to toxicity. Furthermore, the molecule must possess "Pharmacokinetics" (PK) suitable for the intended route of administration. This means it must be absorbed into the systemic circulation, distributed to the target tissue, metabolized at an appropriate rate, and excreted safely.

The challenges in this field are profound. The chemical space of drug-like molecules is estimated at 10^{60}, a vastness that makes exhaustive testing impossible. Moreover, the biological systems they interact with are highly complex and dynamic. Traditional methods often relied on trial-and-error, but the modern era demands a more rational, data-driven approach. Failure in lead optimization often stems from "molecular obesity"—the tendency for compounds to become larger and more lipophilic during the quest for potency, which subsequently ruins their solubility and metabolic stability. The goal of lead optimization is therefore to increase the

probability of clinical success by delivering a molecule that is not just a "hit," but a viable, developable medicine.

3.2 Overview of Lead Optimization

Lead optimization is the strategic process of enhancing the pharmacological properties of a chemical lead to identify a drug candidate. This phase is typically initiated once a lead series has been identified that shows promising activity and a clear path for chemical modification. The core objectives of this stage are to maximize therapeutic efficacy while minimizing adverse effects and ensuring that the compound can be formulated and delivered effectively.

3.2.1 Primary Objectives: Potency, Selectivity, and Drug-likeness

The first objective is the enhancement of potency. For many therapeutic targets, a potency in the range of 1nM to 00nM is required to achieve therapeutic effects without requiring high, potentially toxic, doses. However, potency alone is insufficient. Selectivity ensures that the compound binds to its intended target and not to related proteins. For instance, in kinase inhibition, achieving selectivity across the kinome is essential to prevent unintended interference with cell signaling pathways that are crucial for normal physiological function.

Drug-likeness is a qualitative concept that refers to how closely a compound resembles successful drugs in its physical and chemical properties. This is often benchmarked against "Rule of Five" (Ro5) criteria, which include molecular weight (MW < 500 Da), lipophilicity (Log p < 5), and the number of hydrogen bond donors and acceptors. Compounds that fall outside these ranges are frequently associated with poor oral bioavailability and higher clearance rates.

3.2.2 Key Parameters for Evaluation

To guide the optimization process, a variety of parameters are tracked through experimental assays and computational models. These can be broadly categorized into biological activity, physicochemical properties, and the toxicity profile.

Table 3.1: Variety of parameters are tracked through experimental assays and computational models

Parameter Category	Specific Metrics	Importance in Lead Optimization
Biological Activity	IC_{50}, EC_5, K_i, K_d	Quantifies the strength and nature of

		binding to the primary target and off-targets.
Physicochemical	Log P, Log D_7, Solubility, pKa, TPS A	Dictates the solubility, membrane permeability, and overall developability of the molecule.
Pharmacokinetics	$t_{1/2}$, Clearance (CL), V_{88}, Bioavailability (F)	Predicts how the drug will behave in a living organism over time.
Safety/Toxicity	hERG IC_{50}, CYP inhibition, Genotoxicity	Identifies risks for cardiac arrhythmia, drug-drug interactions, and systemic toxicity.

3.3 Strategies in Lead Optimization

Lead optimization utilizes a diverse toolkit of strategies to navigate the "Design-Make-Test-Analyze" (DMTA) cycle. This cycle is the heart of medicinal chemistry, where structural modifications are designed based on biological feedback, synthesized in the lab, and then tested in various assays to determine the success of the modifications.

3.3.1 Structure–Activity Relationship (SAR)

SAR analysis is the systematic study of how changes in the molecular structure of a compound affect its biological activity. By making subtle alterations—such as adding a methyl group, swapping a chlorine for a fluorine, or shifting a nitrogen atom in a ring—chemists can deduce the structural requirements for binding.

3.3.2 Functional Group Modification and Pharmacophore Modeling

The identification of the pharmacophore—the essential arrangement of functional groups and their spatial orientation—is a key outcome of SAR studies. Modification of functional groups allows chemists to probe the binding pocket of the target. For instance, increasing the bulk of a side chain might improve binding if there is an unoccupied hydrophobic pocket, but it might destroy activity if it causes steric clash.

3.3.3 Bioisosterism in Lead Optimization

Bioisosterism is a powerful strategy where a functional group is replaced by another group that has similar physical or chemical properties but offers better metabolic stability or fewer side

effects. Bioisosteres are categorized into "classical" (atoms or groups with similar size/electronics, like F for H or CH_3 for NH_2) and "non-classical" (groups that look different but fulfill the same biological role, like a tetrazole ring for a carboxylic acid).

Table 3.2: Bioisoterism comparison

Parent Group	**Bioisosteric Replacement**	**Common Rationale**
Carboxylic Acid	1H-Tetrazo	Maintains acidity (pKa =) while increasing lipophilicity and potentially reducing metabolic clearance.
Amide	1,2,4-Triazo	Mimics the polarity and geometry of the amide but is much more metabolically stable against proteases.
Benzene	Thiophene / Pyridine	Maintains the planar, aromatic character while introducing a heteroatom for additional H-bonding or improved solubility.
Hydroxyl ($-OH$)	Chlorine ($-Cl$)	Similar size and electronegativity; often used to block metabolic "soft spots" where the $-OH$ might be glucuronidated.

3.4 Structure-Based Drug Design (SBDD)

SBDD relies on the 3D structure of the target protein, typically resolved via X-ray crystallography or Cryo-EM. By visualizing the ligand-protein complex, medicinal chemists can identify specific interactions such as hydrogen bonds, **π-π** stacking, and hydrophobic contacts.

3.4.1 Molecular Docking and Binding Site Interactions

Molecular docking algorithms predict how a molecule will orient itself within a binding site and calculate a "docking score" that estimates binding energy. In lead optimization, docking is used to screen virtual libraries of analogues, prioritizing those that fit perfectly into the pocket. This is particularly useful for optimizing the "Linker" between two fragments or extending a lead into a newly discovered sub-pocket.

3.5 Ligand-Based Drug Design (LBDD)

When the protein structure is unavailable, LBDD uses the known properties of active and inactive compounds to guide optimization. Quantitative Structure-Activity Relationship (QSAR) models correlate structural descriptors (like molecular weight or electrostatic surface area) with biological activity using mathematical equations.

3.5.1 Similarity Searching and Pharmacophore Mapping

LBDD also involves similarity searching, where new scaffolds are identified that share the same pharmacophoric map (the 3D arrangement of H-bond donors, acceptors, and hydrophobic regions) as the original lead but with better drug-like properties.

3.6 Optimization of Physicochemical Properties

The physicochemical profile of a lead often determines whether it will succeed *in vivo*.

- **Lipophilicity (**Log P**):** While high lipophilicity can improve potency by increasing binding affinity for hydrophobic pockets, it is often inversely correlated with solubility and directly correlated with toxicity and metabolic clearance.
- **Solubility:** A drug must be soluble in the gastrointestinal fluids to be absorbed orally. Chemists improve solubility by introducing polar groups, reducing the melting point, or using prodrug strategies.
- **Permeability:** To reach intracellular targets, a drug must traverse the lipid bilayer of the cell membrane. This requires a balance of lipophilicity (to enter the membrane) and hydrophilicity (to remain in solution).
- **Chemical Stability:** Small molecules must remain intact in various environments, including the acidic pH of the stomach and the presence of plasma enzymes.

3.7 ADMET Optimization

The ADMET profile is the primary cause of drug attrition. Lead optimization must address each component:

- **Absorption:** Maximizing bioavailability through improved permeability and solubility while avoiding efflux transporters like P-glycoprotein (P-gp).
- **Distribution:** Controlling plasma protein binding (PPB) and ensuring the drug reaches the intended tissue. For CNS drugs, crossing the blood-brain barrier is a paramount objective.

BIOPRESS

- **Metabolism:** Small molecules are primarily metabolized by Cytochrome P450 (CYP) enzymes in the liver. Lead optimization focuses on reducing "metabolic soft spots" to extend the drug's half-life and avoiding the inhibition of CYP enzymes to prevent drug-drug interactions.
- **Excretion:** Ensuring the drug and its metabolites are cleared appropriately via the kidneys or bile.
- **Toxicity:** Critical concerns include cardiotoxicity mediated by the hERG potassium channel. Blockade of hERG can lead to QT interval prolongation and sudden cardiac death. Strategies to reduce hERG liability include reducing basicity, introducing acidic groups, or using "zwitterionic" designs.

3.8 Prodrug Strategies

Prodrugs are pharmacologically inactive compounds that undergo enzymatic or chemical transformation in the body to release the active drug. This is a common strategy to overcome poor solubility, low bioavailability, or high systemic toxicity.

- **Carrier-linked Prodrugs:** The active drug is covalently attached to a promoiety (e.g., an ester or phosphate group) that improves absorption and is then cleaved *in vivo*.
- **Bio precursors:** These are structurally simple modifications that require metabolic activation (e.g., oxidation or reduction) to form the active drug.

3.9 Scaffold Hopping

Scaffold hopping involves the replacement of the core structure of a lead molecule while maintaining its biological activity. This is used to improve novelty, circumvent patents, or significantly alter the molecule's properties without losing its target-binding capability. A famous example is the transition from the rigid morphine molecule to the more flexible and safer tramadol via "ring opening" scaffold hopping.

3.10 Fragment-Based Drug Design (FBDD)

FBDD starts with very small molecules ("fragments," MW < 300 Da) that bind weakly to the target. These hits are then grown or linked together to form a potent lead. FBDD is particularly effective for challenging targets like protein-protein interactions (PPI_8), where traditional HTS often fails to find suitable starting points.

3.11 Multi-Parameter Optimization (MPO)

MPO is a strategy that uses mathematical scoring functions to balance competing objectives. The Central Nervous System (CNS) MPO score is a widely used example, combining six parameters (C Log P, C Log D, MW, TPSA, HBD, and basic pKa) into a single score from 0 to 6. A score above 4 generally indicates a higher probability of successful drug-likeness for brain-penetrant compounds.

3.12 Role of Computational Tools and AI

The integration of artificial intelligence and machine learning is perhaps the most significant trend in modern lead optimization. These tools enable the rapid navigation of the vast chemical space, predicting the outcomes of modifications before they are synthesized in the lab.

3.13 In silico Screening and Machine Learning

In silico screening allows researchers to evaluate millions of virtual compounds for their likelihood of binding to a target. Machine learning models, trained on historical data, can predict ADMET properties with high accuracy, identifying potential safety issues early in the lead optimization process. For example, deep learning architectures such as Graph Neural Networks (GNNs) can represent molecular structures as graphs, capturing complex relationships between atoms and bonds to predict biological activity and toxicity.

3.14 De Novo Drug Design and Generative Models

Generative AI models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), can "invent" entirely new molecular structures that meet a specific set of parameters. This *de novo* approach bypasses the limitations of existing chemical libraries and can discover novel scaffolds that a human chemist might not consider.

Case Study: Insilico Medicine's ISM001-055

Insilico Medicine recently utilized its AI platform, Pharma.AI, to identify a novel target (TNIK) for idiopathic pulmonary fibrosis (IPF) and design a small-molecule inhibitor, ISM001-055. The entire process from target identification to the nomination of a preclinical candidate took only 18 months and required the synthesis of fewer than 80 compounds, a massive reduction in time and cost compared to traditional industry benchmarks.

BIOPRESS

3.15 Experimental Approaches

Despite the rise of computational tools, experimental validation remains essential.

- **In vitro Assays:** Used to measure target binding affinity (K_d), inhibition potency (IC_{50}), and cellular efficacy. Techniques like Surface Plasmon Resonance (SPR) and Bio-Layer Interferometry (BLI) provide real-time kinetic data.
- **High-Throughput Screening (HTS):** Allows for the rapid testing of thousands of compounds in a short period, although it is typically used earlier in hit identification than in the deep optimization phase.
- **In vivo Pharmacokinetics:** Conducted in animal models to understand the drug's absorption, distribution, half-life, and clearance in a whole organism.
- **Toxicological Evaluation:** Includes both *in vitro* safety screens (like the Ames test for mutagenicity) and *in vivo* toxicity studies to determine the safe dosage range and identify target organs for toxicity.

3.16 Case Studies

3.16.1 Venetoclax: A Triumph of Fragment-Based Drug Discovery

Venetoclax (Venclexta) is a first-in-class inhibitor of Bcl, a protein that prevents cell death (apoptosis) and is overexpressed in many cancers. Targeting $Bcl\text{-}2$ was long considered impossible because the binding site is a large, flat surface involved in protein-protein interactions. Using FBDD, researchers identified small fragments that bound to sub-pockets of the surface. Through structure-guided optimization and fragment linking, they developed ABT-737, which was potent but had poor oral bioavailability. Further optimization led to Venetoclax, which is orally bioavailable and highly selective, demonstrating how FBDD can tackle "undruggable" targets.

3.16.2 Sildenafil to Vardenafil: Scaffold Hopping in PDE5 Inhibition

Pfizer's Sildenafil (Viagra) was the first successful $PDE5$ inhibitor for erectile dysfunction. Bayer later developed Vardenafil (Levitra) by performing a scaffold hop—swapping a nitrogen atom's position in the core pyrazolo pyrimidinone ring system. This modification resulted in a molecule with similar biological activity but distinct chemical properties, allowing for a novel patent and competitive positioning in the market.

BIOPRESS

3.16.3 Morphine to Tramadol: Structural Simplification

Morphine is a powerful but dangerous analgesic with a complex pentacyclic structure. Through "ring opening" scaffold hopping, chemists developed Tramadol. By breaking several bonds in the morphine core, they created a simpler, more flexible molecule that retained its analgesic properties while significantly reducing the risk of respiratory depression and addiction.

Table 3.3: Case study

Case Study	Target	Original Lead	Key Modification	Outcome
Venetoclax	$Bcl\text{-}2$	Fragment Hits	Fragment Linking & SBDD	First-in-class, selective oral cancer drug.
Vardenafil	$PDE5$	Sildenafil	Heterocycle Replacement	Successful second-generation drug with novel IP.
Tramadol	μ-opioid	Morphine	Ring Opening / Simplification	Improved safety profile and oral absorption.
ISM001-055	$TNIK$	*AI-generated*	Generative Design	Rapid progression to clinical trials (18 months).

3.17 Challenges and Limitations

Lead optimization is often referred to as the "bottleneck" of drug discovery because of the inherent difficulty in optimizing multiple parameters simultaneously.

- **Potency vs. Toxicity:** The "Lipophilicity Trap" is a common challenge where increasing potency by making a molecule more lipophilic simultaneously increases its affinity for off-targets like hERG or metabolic enzymes, leading to toxicity.
- **Poor Translation:** Success in animal models or *in vitro* assays often fails to translate to humans due to differences in metabolic pathways or target biology.
- **Time and Cost:** Traditional lead optimization can take 3-5 years and cost hundreds of millions of dollars, yet most candidates still fail in clinical trials.

BIOPRESS

- **Resistance:** In oncology and infectious diseases, targets can mutate to become resistant to small molecules, requiring constant "retro-optimization" to find new ways to bind to the mutant forms.

3.18 Emerging Trends

The landscape of lead optimization is shifting toward more sophisticated and sustainable approaches.

- **AI-Driven Optimization:** The use of generative models and automated DMTA cycles is set to become the standard, significantly reducing the time to clinical entry.
- **Microfluidics and Organ-on-a-Chip:** These technologies provide a more human-like environment for testing ADMET and toxicity, potentially reducing the need for animal models.
- **Personalized Medicine:** Lead optimization is increasingly focused on designing drugs for specific genetic subsets of a disease, identified through "omics" data.
- **Green Chemistry:** There is a growing emphasis on making drug synthesis more environmentally friendly by reducing waste, using renewable feedstocks, and replacing toxic solvents with safer alternatives like ethanol or water.

3.19 Future Perspectives

The future of small molecule development will likely see the full integration of AI and experimental biology. We can expect "closed-loop" systems where AI-driven robots autonomously design, synthesize, and test new molecules in a continuous cycle. Furthermore, the rise of "Targeted Protein Degradation" (TPD) and PROTACs represents a paradigm shift where small molecules are used not just to inhibit a protein, but to destroy it entirely, opening up vast numbers of previously "undruggable" targets. As precision medicine matures, lead optimization will become more bespoke, with molecules designed for specific patient profiles from the outset.

3.20 Conclusion

Lead optimization is a multidisciplinary and iterative process that is fundamental to the success of small molecule development. By combining the intuition of medicinal chemists with the predictive power of computational tools and the precision of experimental assays, researchers can transform weak hits into potent, selective, and drug-like candidates. While challenges such

as clinical attrition and rising R&D costs remain significant, innovations in AI, fragment-based design, and ADMET modeling are providing new pathways to overcome these hurdles. The ultimate success of lead optimization is measured not just in binding affinity, but in the delivery of transformative medicines that improve human health.

3.21 References

1. New strategies to enhance the efficiency and precision of drug ..., https://pmc.ncbi.nlm.nih.gov/articles/PMC11850385/
2. Lead Optimization in Drug Discovery | Danaher Life Sciences, https://lifesciences.danaher.com/us/en/library/lead-optimization-drug-discovery-guide.html
3. Drug Discovery → Lead Optimization - Developing Medicines, https://drugdevelopment.web.unc.edu/drug-discovery-%E2%86%92-lead-optimization/
4. Advances in Multi-parameter Optimisation Methods for de Novo Drug Design | Optibrium, https://www.optibrium.com/wp-content/uploads/2020/12/Advances-in-MPO-for-de-Novo-Drug-Design-preprint.pdf
5. Progress of AI-Driven Drug–Target Interaction Prediction and Lead Optimization - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC12563295/
6. Hit-to-Lead process | Drug Discovery - Oncodesign Services, https://www.oncodesign-services.com/drive-idds/hit-to-lead/
7. Drug development hit and lead optimization - Pion Inc, https://www.pion-inc.com/blog/hits-and-lead-optimization-in-drug-development
8. Lead Optimization in Drug Discovery: Process, Strategies & Tools | Chemspace, https://chem-space.com/news/chemspace-lead-otimization
9. Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery | ACS Chemical Neuroscience - ACS Publications, https://pubs.acs.org/doi/10.1021/acschemneuro.6b00029
10. Multi-parameter Optimization: Identifying High Quality Compounds With a Balance of Properties - PubMed, https://pubmed.ncbi.nlm.nih.gov/22316157/
11. Prodrugs for Improved Drug Delivery: Lessons Learned from Recently Developed and Marketed Products - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC7692606/
12. Structural simplification: an efficient strategy in lead optimization - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC6804494/
13. Modern Lead Optimization: A Journey through Science and Innovation - PharmaFeatures, https://pharmafeatures.com/modern-lead-optimization-a-journey-through-science-and-innovation/

BIOPRESS

14. What Happens in Successful Optimizations? A Survey of 2018–2024 Literature | Journal of Medicinal Chemistry - ACS Publications, https://pubs.acs.org/doi/10.1021/acs.jmedchem.5c03171

15. AI/ML-Enabled Drug Discovery Part 1 | Discovery on Target | Sept 22 - 25, 2025, https://www.discoveryontarget.com/ai-ml-enabled-drug-discovery-part1

16. Optimization of hERG and Pharmacokinetic Properties for Basic Dihydro-8H-purin-8-one Inhibitors of DNA-PK - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC9377022/

17. Recent developments in anticancer kinase inhibitors based on the pyrazolo[3,4-d]pyrimidine scaffold - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC7652001/

18. Multi-Parameter Optimization: Identifying high quality compounds with a balance of properties - Optibrium, https://optibrium.com/downloads/MPO_review_preprint.pdf

19. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC3368654/

20. Lipophilic efficiency - GARDP Revive, https://revive.gardp.org/resource/lipophilic-efficiency/?cf=encyclopaedia

21. Structural Modification in Anesthetic Drug Development for Prodrugs and Soft Drugs - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC9283706/

22. Machine Learning in Drug Discovery Market 2025 Trends - Towards Healthcare, https://www.towardshealthcare.com/insights/machine-learning-in-drug-discovery-market-sizing

23. Dealing with hERG liabilities early: diverse approaches to an important goal in drug development – PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC2823348/

24. Lead compound optimization strategy (5) - Reducing the hERG cardiac toxicity in drug development - ResearchGate, https://www.researchgate.net/publication/317828735_Lead_compound_optimization_strategy_5_-_Reducing_the_hERG_cardiac_toxicity_in_drug_development

25. Back to the Future of Lead Optimization: Benchmarking Compound Prioritization Strategies | ChemRxiv, https://chemrxiv.org/doi/10.26434/chemrxiv-2025-5kc10

26. Hit to Lead Optimization in Drug Discovery - Excelra, https://www.excelra.com/glossary/hit-to-lead-optimization/

27. Acid Bioisosteres - Cambridge MedChem Consulting, https://cambridgemedchemconsulting.com/acid-bioisosteres/

28. Bioisosteres for Drug Hunters: Part 1 - Background, Carboxylic Acids, and Amides, https://drughunter.com/resource/bioisosteres-for-drug-hunters-part-1-background-carboxylic-acids-and-amides

BIOPRESS

29. Drug Modifications to Improve Stability – An Introduction to Medicinal Chemistry & Molecular Recognition - eCampusOntario Pressbooks, https://ecampusontario.pressbooks.pub/medicinalchemistry/chapter/drug-modifications-to-improve-stability/

30. Scaffold Hopping in Drug Discovery - NIPER, https://www.niper.gov.in/sites/default/files/inline-files/Vol16No5_Page94_99.pdf

31. Medicinal Chemistry of hERG Optimizations: Highlights and Hang-Ups - ACS Publications, https://pubs.acs.org/doi/10.1021/jm060379l

32. Efficient Drug Lead Discovery and Optimization - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC2727934/

33. Bridging traditional and contemporary approaches in computational medicinal chemistry: opportunities for innovation in drug discovery - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12641420/

34. Deep Lead Optimization: Leveraging Generative AI for Structural ..., https://pubs.acs.org/doi/10.1021/jacs.4c11686

35. 9905 PDFs | Review articles in LEAD OPTIMIZATION - ResearchGate, https://www.researchgate.net/topic/Lead-Optimization/publications

36. Artificial Intelligence and Machine Learning in Drug Discovery: Trends, Techniques, Challenges, and Future Prospects - International Journal of Pharmaceutical Sciences, https://www.ijpsjournal.com/article/Artificial+Intelligence+and+Machine+Learning+in+Drug+Discovery+Trends+Techniques+Challenges+and+Future+Prospects

37. The Prodrug Approach: A Successful Tool for Improving Drug Solubility - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC6273601/

38. Prodrug Approach as a Strategy to Enhance Drug Permeability - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC11946379/

39. Challenges and Strategies in Prodrug Design: A Comprehensive Review, https://www.sciensage.info/index.php/JASR/article/download/2430/1715

40. Go-Go Gone: Smart hERG Optimization Strategies - Drug Hunter, https://drughunter.com/commentary/go-go-gone-smart-herg-optimization-strategies

41. Prodrugs in Modern Medicine: Strategies, Applications, and Future Directions - SciRP.org, https://www.scirp.org/journal/paperinformation?paperid=145302

42. Scaffold hopping, https://elearning.uniroma1.it/pluginfile.php/1384573/mod_folder/content/0/Sezione%205.0/5.4/5.4.1.2.10.0.1.2004.Scaffold%20hopping.pdf?forcedownload=1

BIOPRESS

43. Classification of Scaffold Hopping Approaches - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC3328312/
44. Fragment-based drug discovery: A graphical review - PubMed, https://pubmed.ncbi.nlm.nih.gov/41473206/
45. A Role for Fragment-Based Drug Design in Developing Novel Lead Compounds for Central Nervous System Targets - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC4566055/
46. Fragment-Based Drug Design: From Then until Now, and Toward the Future | Journal of Medicinal Chemistry - ACS Publications, https://pubs.acs.org/doi/10.1021/acs.jmedchem.5c00424
47. Enabling synthesis in fragment-based drug discovery (FBDD): microscale high-throughput optimisation of the medicinal chemist's toolbox reactions - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC10718589/
48. Probabilistic Approach to Generating MPOs and Its Application as a Scoring Function for CNS Drugs - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC4716604/
49. Optimizing Lead Compounds: The Role of Artificial Intelligence in Drug Discovery, https://www.preprints.org/manuscript/202404.0055/v1
50. Expediting hit-to-lead progression in drug discovery through reaction prediction and multi-objective molecular optimization | ChemRxiv, https://chemrxiv.org/doi/10.26434/chemrxiv-2025-0lxhw
51. Insilico Pharma.AI: AI Drug Discovery Platform Analysis | IntuitionLabs, https://intuitionlabs.ai/articles/insilico-pharma-ai-drug-discovery
52. Case Study: Insilico's Transformation, https://insilico.com/casestudy
53. Insilico Medicine Reports Positive Phase IIa Results for ISM001-055, a Novel First-in-Class Drug Treatment for Idiopathic Pulmonary Fibrosis (IPF) Designed Using Generative AI - PR Newswire, https://www.prnewswire.com/news-releases/insilico-medicine-reports-positive-phase-iia-results-for-ism001-055-a-novel-first-in-class-drug-treatment-for-idiopathic-pulmonary-fibrosis-ipf-designed-using-generative-ai-302251732.html
54. Discovery and Development of Venetoclax, a Selective Antagonist of BCL-2 - ResearchGate, https://www.researchgate.net/publication/335605900_Discovery_and_Development_of_Venetoclax_a_Selective_Antagonist_of_BCL-2
55. Hit to Lead Assays: Accelerating Early Drug Discovery - BellBrook Labs, https://bellbrooklabs.com/hit-to-lead-assays/
56. Application and development of Organ-on-a-Chip technology in cancer therapy - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC12446035/
57. Organ-on-a-Chip Applications in Microfluidic Platforms - MDPI, https://www.mdpi.com/2072-666X/16/2/201

BIOPRESS

58. Integrating In Silico Tools with Organ-on-a-Chip to advance ADME studies, https://cn-bio.com/integrating-in-silico-tools-with-organ-on-a-chip-to-advance-adme-studies/

59. Green chemistry: Six key trends to watch - CAS, https://www.cas.org/resources/cas-insights/green-chemistry-trends

60. Embracing Sustainable Processes in the Pharmaceutical Industry with Green Chemistry and Engineering - Scifiniti, https://www.scifiniti.com/3079-1421/1/2025.0013

61. The Role of Green Chemistry in Generic Drug Development: A Sustainable Approach to Medicine - DrugPatentWatch – Transform Data into Market Domination, https://www.drugpatentwatch.com/blog/the-role-of-green-chemistry-in-generic-drug-development-a-sustainable-approach-to-medicine/

62. The Future of Green Chemistry: Evolution and Recent Trends in Deep Eutectic Solvents Research - MDPI, https://www.mdpi.com/2076-3417/16/2/654

BIOPRESS

Chapter 4

ADMET Prediction and Pharmacokinetic Modeling

Dr. K. P. Beena

Associate Professor, Department Of Pharmaceutical Chemistry,

Sri Ramakrishna Institute of Paramedical Sciences, College Of Pharmacy,

Coimbatore, Tamil Nadu, India

Abstract

ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) prediction and pharmacokinetic modeling play a pivotal role in modern drug discovery and development by enabling early assessment of a compound's biological and safety profile. *In silico* ADMET tools facilitate rapid screening of drug candidates, helping to identify potential limitations such as poor bioavailability, metabolic instability, or toxicity risks before costly experimental studies are undertaken. These computational approaches rely on quantitative structure–activity relationships (QSAR), machine learning algorithms, and large chemical databases to generate reliable predictions.

Pharmacokinetic modeling complements ADMET prediction by providing a quantitative framework to describe drug absorption, distribution, metabolism, and elimination within the body. Techniques such as compartmental modeling and physiologically based pharmacokinetic (PBPK) modeling allow for simulation of drug concentration–time profiles and prediction of dose-response relationships. Together, these approaches support rational drug design, dose optimization, and risk assessment.

The integration of ADMET prediction with pharmacokinetic modeling enhances decision-making, reduces drug attrition rates, and accelerates the development process. As computational technologies continue to evolve, these methods are expected to become increasingly accurate and widely applicable, contributing significantly to the design of safer, more effective therapeutic agents and advancing personalized medicine.

Key words: Absorption, Distribution, Metabolism, Excretion, Pharmacokinetics

4.1 Introduction

Chemoinformatics is an area of study based on the chemical information for the design and development of safe and effective drugs using computer technologies. Computational techniques are the computer tools used for solving the problems by step-wise, repeated and iterative methods, which would otherwise be tedious or unsolvable by manual calculations.

A series of events occurring starts from the point of drug administration until the drug gets eliminated from the body. Liberation of the drug from a dosage form is a key process in

bioequivalence studies of oral products or for parenteral formulation, wherein the rate of drug release from the formulation is determined. In the drug treatment, drug administration can be optimized to elicit the desired clinical response.

Drug disposition includes DME, while pharmacokinetics includes ADME. The preclinical drug development assesses the safety and effective dose of drugs. These aspects do not provide adequate information on the drug bioavailability. In other words, the optimization of absorption, distribution, metabolism, excretion and toxicity (ADMET) became essential. ADME are studied using in vitro biochemical methods and in vivo animal models. Statistical tools and algorithms are used to correlate the chemical structural information with the biological transport and processes. Such models support to develop computer simulations.

4.1.1 Drug disposition

Drug disposition refers to the quantification of the time course of changes in the concentration of drug or metabolite in the processes occurring after the absorption of the drug i.e. distribution, metabolism and excretion (DME). The computational approaches are described individually in each process. Though ADME is frequently referred, DME is evaluated initially, because these are inherent to the body, intrinsic properties of the drug, i.e. do not change.

4.1.2 Pharmacokinetics

Pharmacokinetics is concerned with the quantification of the time course of changes in the concentration of drug or metabolite in the processes, such as absorption, distribution, metabolism and excretion (ADME).

Pharmacokinetics deals with the mathematical description of drug transfer through the body tissues or fluids. The pharmacokinetic parameters are evaluated through experimentation, in vitro models using isolated organs and in vivo models employing healthy human volunteers and/or patients with a specific disease state.

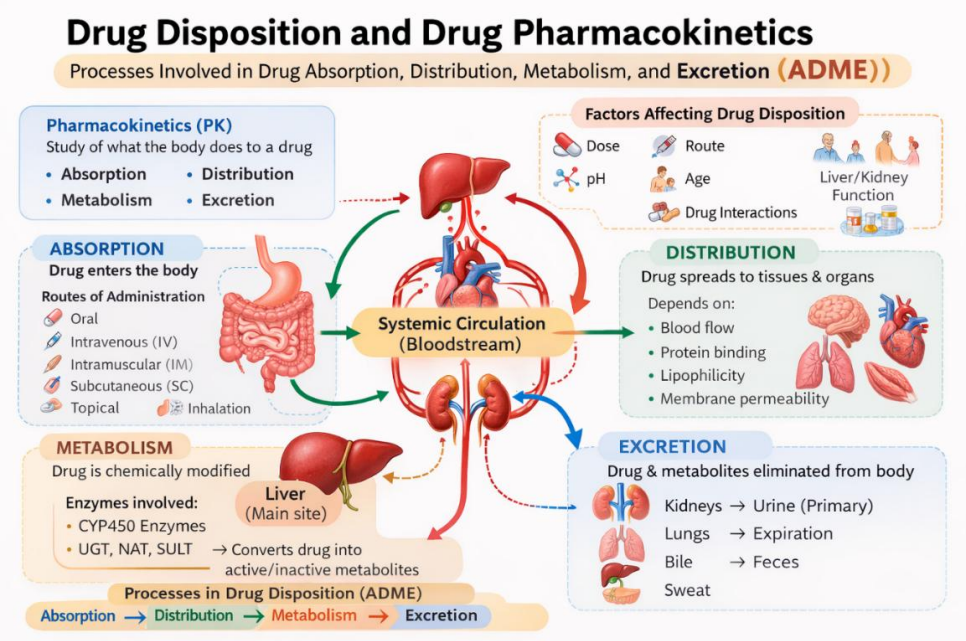


Figure 4.1 Drug Disposition and Drug Pharmacokinetics

4.1.3 Molecular descriptors-Modeling

The molecular descriptor is defined as the final result of logic and mathematical procedures, which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result or some standardized experiment.

Molecular descriptors are expected to reflect the interactions of the drugs with the targets, resulting in eliciting the biological response. Hence it is equally important that the knowledge of the target is the pre-requisite to initiate the correlations.

The descriptors are allowed to fit into 2D or 3D architecture of the enzymes. Molecular descriptors are closely related to the physical properties and easy to comprehend. Examples are molecular weight, solubility, permeability etc.

Molecular descriptors are divided into two main categories:

- Experimental measurements (physico-chemical properties, eg. Log P, molar refractivity, dipole moment and polarizability)
- Theoretical molecular descriptors (interaction energies-quantum mechanical concepts)

The selection of a descriptor must be appropriate to relate the right process. It may be necessary to use more than one statistical method to compare the results for identifying the best possible approach. Several computational approaches are described and related commercial software are developed for modeling, but these may lack the desired specificity. Sometimes, these may not be within the chemical space of the company. In silico models are predictive, when applied to the same chemical space, i.e. the training compounds. Hence, the utility of the model to a different chemical space is a limitation

Table 4.1: In silico descriptors for ADMET

Physicochemical properties	Aqueous solubility (Log S)
	Permeability-Lipophilicity (Log P)
	Ionization (pKa) and Log D
	Polar surface area (PSA)
Physiological parameters	Human intestinal absorption (HIA)
	Plasma protein binding (PPB)
	Perfusion rate (PR)
	Volume of distribution (V_1)
Intrinsic parameters	Age
	Organ size
	Pregnancy
	Genetic variability
Extrinsic parameters	Formulation type

	Drug interactions

4.2 Objectives of computational models in ADME

The drug formulation requires the optimization of bioprocesses (ADMET). The clinical trial procedures are very expensive and also include the expenditure on the clinical failures. Computer aided drug development (CADD) is a useful tool for evaluating ADME-Tox processes.

(a) The relationship between the structure and drug utilization pathways (in the body) can be established

(b) The electronic effects of the chemical structure are implicated for predicting the ADMET profiles

(c) Hydrophobic, hydrogen bonding (donor/acceptor) and ionizable features are implicated through K_i, IC_{50} and percent inhibition properties

(d) The site of metabolism and possible metabolites can be predicted

4.2.1 Advantages of Computational models

Computational techniques are developed for modeling which supports to predict the drug pharmacokinetics. Several molecular descriptors are implicated as given below

(a) In silico computational methods are attempted initially for optimization of a process, Eg. Solubility, partitioning, diffusion coefficient and binding constants

(b) High throughput in vitro ADMET property screening assays are developed. Eg. Caco-2 and MDCK cell monolayers

(c) The analysis of transporters is successful, because the substrates and inhibitors are made available, Eg. Multidrug resistance associated protein family (MRP), breast cancer resistance protein and organic anion transporters (OAT) are a few

(d) Active transport can be modelled using software, such as GastroPlus^{TM}, PK-Sim and ADME/Tox WEB. Several pharmacophore models are developed.

4.2.2 Limitations of Computational Models

The ADMET modeling has several limitations

(a) The quality of data is low. It is the weakest link in the analysis

(b) The practical applications are low, though empirical relationships are established

(c) The number of transporters is many, but a few are characterized

(d) Several transporters are of general nature, not exclusive for an individual process of ADMET

(e) The number of models is many

4.3 Drug absorption- Computational models

Computational models permit the analysis of complex transport phenomena. It is also possible to predict the intestinal absorption based on the chemical structure. A few physiochemical properties, such as log P (partition coefficient) and polar surface area (PSA) are related to hydrogen bonding potential.

Almost a perfect correlation between the partition coefficient and extent of absorption was observed for different barbiturates. Sophisticated and accurate in silico tools used for the computations are: QMPRPlus (Quantitative molecular permability relationship) and GastroPlus, etc.

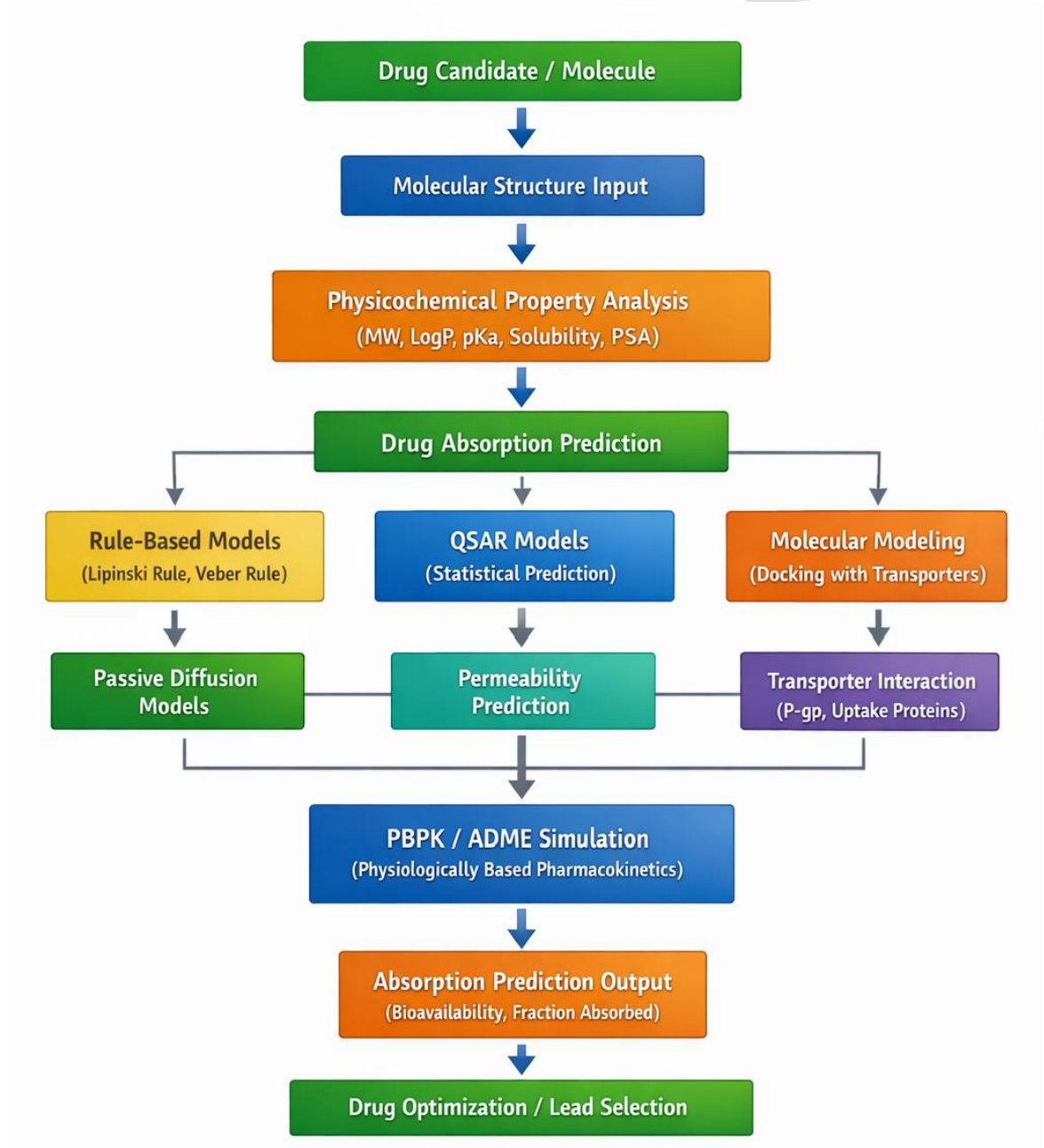


Figure 4.2: Computational methods for predicting drug absorption in GIT

4.4 In silico techniques based on physicochemical properties

4.4.1 Solubility:

For drug like molecules, solvent-solute interactions are the major determinants of solubility

- Log P is the simplest estimation and uses program, cLogP, based on fragment approach
- The contribution of solute crystal lattice energy is included by amending log P values for accurate prediction
- Log P values (n-octanol/water) and melting points can be used to estimate solubility indirectly
- Multivariate analysis method is used to identify the correlations between molecular descriptors and solubility

These efforts considered the dissolution process, but ignored the underlying physiological processes. Selection of field descriptors that reflect the physiological process becomes more successful. The target property is the logarithm of solubility (log S). The solubility of a compound can be estimated using in silico modeling, before synthesizing the compound.

Two methods are available.

- Approach using physiological process
- Approach of empirical means

The models are trained and verified with the AQUASOL and PhysProf databases.

4.4.2 Intestinal permeability

Intestinal permeation means drug absorption. It is a complex process that is difficult to predict solely based on molecular mechanisms. This process involves both passive diffusion and active transport. The in vivo drug absorption is simulated using models such as

- ❖ In vitro membrane permeation of Caco-2
- ❖ Madin-Darby canine kidney (MDCK) cells
- ❖ Parallel artificial membrane permeability assay (PAMPA)

4.4.3 Ionization of drugs

The ionization will affect the solubility and permeability, which in turn influence drug absorption. The GIT pH and the dissociation constant of the drug determine the extent of ionization. The pKa values of the substances can be obtained from the chemical structure using software, CSS pKa, Pallas/pKalc and ACD/pKa.

4.5 QSAR (Quantitative Structure Activity relationship)

Statistical QSAR models are developed using the PAMPA and Caco-2 cell assay results. Hence, human drug absorption can be predicted. The results can be extended for MDCK cell permeability predictions and transporter interactions.

BIOPRESS

- Human oral absorption prediction – vande Watermed QSAR equation
    - Log Papp = 0.08 (±0.002) MW-0.043 (±0.008) PSA-5.165 (±0.605)
- Human intestinal absorption (%HIA) -Model equation
    - % HIA = f (Log $D_{7.4}$, molecular size, Hydrogen bond capacity)
- Human oral bioavailability-Halogram QSAR (HQSAR) model. For early ADMET evaluation
- Human intestinal permeability-QSAR models. ASMS (aqueous solubility based on the molecular surface) and ASMS-LOGP (ASMS using cLogP)

4.6 Transporters for absorption modeling

Transporters mediate drug absorption. In vitro cell assay, such as PAMPA assay, Caco-2 cell assays showed good correlations. Caco-2 cell assay results can be used in predicting transporters and CYP3A drug metabolizing enzymes.

4.6.1 Human intestinal peptide transporter (hPEPT1)

The hPEPT1 is a proton coupled, low affinity, active oligopeptide transport system. This pharmacophore model is applied to screen the CMC database with over 8000 drug like molecules.

The features recognized for hPEPT1 transporter are:

- Two hydrophobic features
- One hydrogen bond donor
- One hydrogen bond acceptor
- One negative ionizable feature

4.6.2 Apical sodium-dependent bile acid transporter (ASBT), influx

The ASBT is expressed on the apical membrane of intestinal epithelia and cholangiocytes. ASBT assists in the absorption of bile acids and has a pivotal role in cholesterol metabolism.

The features recognised for ASBT transporter are:

- One hydrogen bond donor
- One hydrogen bond acceptor
- One negative charge
- Three hydrophobic centres

4.6.3 Nucleoside transporters

Nucleoside transporters are:

Concentrative nucleoside transporters

Equilibrative nucleoside transporters

Several transporters for predicting the drug absorption are

- P-gp (ABCB-1)

BIOPRESS

- BCRP/MXR (ABCG2)
- MRP-1 (ABCC1)
- MRP-2 (ABCC1)
- PEPT-1 (SLC15A)
- OCTN-1(SLC22A)
- OCTN-2(SLC22A)
- OCT-1/OCT-2(SLC22A)
- PMAT(SLC29)
- OATP2B1(SLC)
- OATP1A2(SLCO)
- MCT-1(SLC16)

4.7 Drug distribution- Computational Models

Computational descriptors are not yet developed, due to lack of drug interaction with living cells. Predictions are attempted for neutral and basic drugs using physicochemical properties: solubility, partition coefficient (n-octanol/water) and ionization

Transporter systems-Computational models

BCRP pharmacophore model

This model is developed from GF120918, Ko143, Nelfinavir and Nicardipine

- One amine group is bonded to a heterocyclic carbon
- Fused heterocyclic rings and two substituents on a carbocyclic ring of the fused heterocyclic ring
- Three hydrogen bond acceptors (HBAs)
- Three hydrophobic features

BCRP 3D-QSAR model

This model was generated by analyzing the activity of twenty-five flavonoid analogues. The features are:

- 2,4 – double bond in ring C
- Hydroxylation at position 5

P-Glycoprotein (P-gp)

It mediates the efflux of endogenous compounds and drugs out of cells, through ATP dependent mechanism.

Blood Brain Barrier- Computational methods

If CNS is the target, the drug should cross the BBB to gain entrance into CNS for targeting. The following methods are used for modeling.

Log blood/brain (log BB) modeling

BIOPRESS

It measures the drug partitioning between blood and brain tissue. It indirectly indicates BBB permeability and characteristic of the drug. It will not discriminate between free or plasma bound drug. Log D parameter is used for the predications.

Norinder rule of two

This rule is applied to drug transport through BBB. The molecules with the sum of nitrogen and oxygen atoms less than five will tend to permeate through BBB. The positive number obtained by subtracting the sum of N + O from log P is also indicative of better permeability through BBB.

Ionization pka

Positive charge at pH 7 to 8 favours the brain permeability. Positively charged drugs can permeate the brain with a greater fold, than those predicted by pH partition hypothesis. Quarternary nitrogen compounds with logD value between 1-3 favour the brain entry of CNS drugs.

Polar surface area (PSA)

Polar surface area has a distinct relationship with blood brain partitioning. Lower PSA is considered with the ability to cross BBB. Log BB reflects the total drug concentration in brain.

Log P (o/w): It reflects the protein binding, which limits the drug transport across BBB

Log P (chloroform/water): It played a role in the partitioning process in non-polar regions of the brain

Efflux transporters modeling

Along with complex tight junctions in BBB, efflux transporters and metabolic enzymes are also present. Five types of drug efflux transporters are reported

- Blood brain barrier choline transporters (BCT)
- Multidrug resistance transporters (MRT)
- Monocarboxylic acid transporters (MCT)
- Organic ion transporters (OAT and OCT)
- P-gp

Blood brain barrier-choline transporter (BCT)

Choline like compounds (positively charged) can easily penetrate through BBB. Binding studies are developed from theoretical and empirical methodologies. Compounds having diverse structures are studied for their in-situ rat brain perfusion. The empirical data is utilized to build QSAR model. The features recognized for this system are

- Three hydrophobic interactions
- One hydrogen bonding interaction around ammonium moiety (positive ion)

Plasma protein binding-*In silico* methods

BIOPRESS

The plasma protein binding (PPB) is an important consideration, while evaluating the effectiveness (unbound) of a drug. Nonlinear regression method was adopted for analysis using experimental human PPB % data of about 300 drugs. Sigmoidal correlation is found between log D (distribution coefficient) and PPB. The model was validated with an external test set of 20 compounds. This method provides a useful approximation. Several limitations are identified.

- The model should not rely on the binding data of one protein
- In fact, composite parameters reflect the interaction of multiple proteins
- Further each protein may have several binding sites complicating the analysis

Thus, the binding is concentration dependent.

4.8 Biotransformation – Computational models

Liver toxicity is one of the major causes for dropping the drug at preclinical and clinical studies. The drug induced hepatotoxicity leads to organ failure and ultimately death. The *in-silico* models based on chemical structure provides a cost-effective approach for countering the hepatotoxicity. Normally, a small homologous set of molecules are tested using enzymes and fewer molecular descriptors are identified. The drug metabolism is considered as a two-step process

- Drug (substrate) -enzyme binding: The drug binds to the enzymes. The binding studies based on QSAR are applied, particularly using CYP450. Lipophilicity, log P and molecular refractivity are the earliest descriptors. Steric, electronic and molecular shape properties are implicated in the analysis of drug-enzyme binding
- The release of metabolite (modified substrate) from the enzyme: The metabolite is released from the enzymes. This is considered as an opposite to binding. Therefore, QSMR (quantitative structure metabolite relationship) models are constructed.

4.8.1 Objectives of computational methods for drug metabolism

Metabolism is just one component contributing to the success for the drug development. The objectives of the computational techniques in drug metabolism are

- To improve the success of molecules selected to clinical trials
- To develop in silico methodologies to uncover the relationship between chemical structure and metabolic activity of compounds
- To explain the complete biotransformation pathways to define activity, toxicity and interactions.

4.8.2 Concept of chemical space and predictions

More than 10^{60} molecules exist theoretically in the universe and constitute chemical space. In that 10^{20}-10^{24} molecules can be synthesized and becomes real chemical space. Among these, 107 molecules have biological details and are known as biological space.

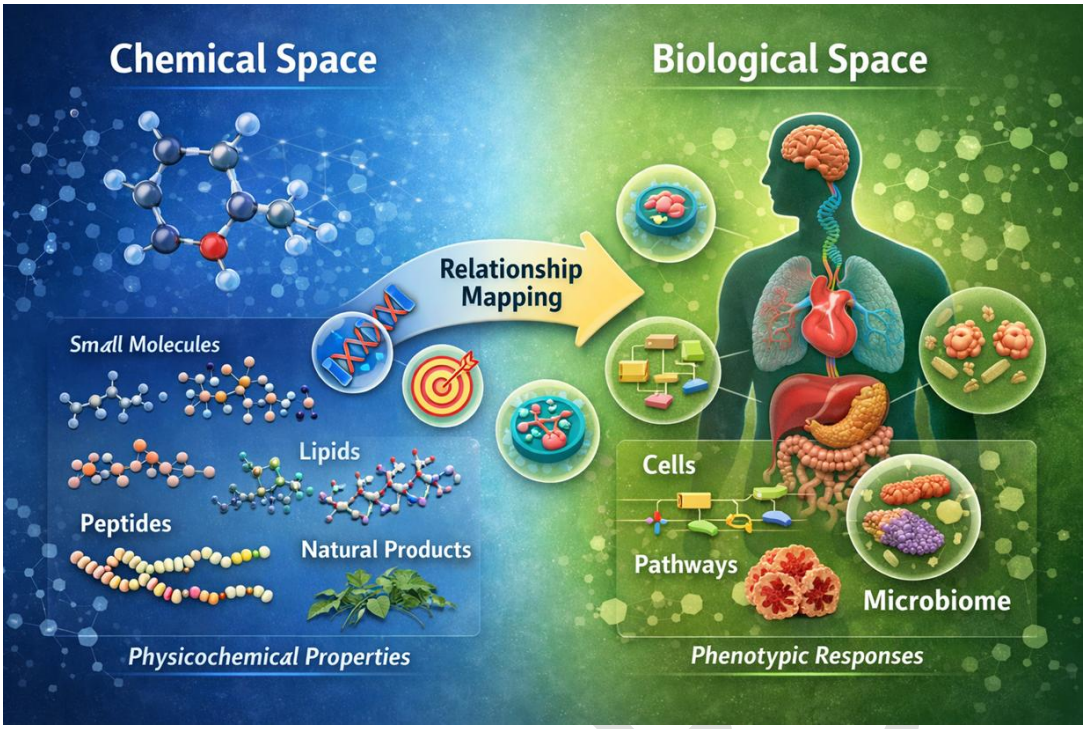


Figure 4.3: Correlation of chemical space with biological space

It has been estimated that 20 to 40% of investigated drugs fail in drug development can be attributed to drug toxicity aspects.

4.8.3 Limitations of computational models of metabolism

The interactions between molecules and enzymes are likely to become more complex. Most of the analysis involves the experimental measurements.

- The metabolic pathways can be different even in close mammalian species
- Metabolism depends on the expression of particular enzymes or disease state
- The applicable chemical space is of the order or 10^{20} to 10^{24} molecules
- The metabolic data are available for mice and rats, which help in differentiating the pathways. However, human metabolism predictions are desirable
- The predictions are prone to errors due to the complexity of biological species differences
- The limitations of animal data include sex differences in xenobiotic metabolism
- The reported Vmax values for the compounds may vary by 2-3 orders of magnitude. This produces unreliable QSAR modeling
- There is a need for developing more consistent datasets for each enzyme
- The rule-based metabolism software may generate many more metabolites. This also reflects the sensitivity of analytical methods at that time

Machine learning algorithms need to be appropriately developed for predicting metabolites under experimental conditions. Computational models are continuously updated at regular intervals.

4.9 Drug Metabolism-Software

The metabolic fingerprint of the libraries of compounds permits through the use of software in screening. Enzyme crystal structure beyond CYP2C9 and CYP3A4 are needed to improve the homology models. One approach is calculating the Tanimotto similarity score using molecular descriptors for both molecules in the model training set.

Table 4.2: Commercial software available for computational techniques

Software	Functions	Website
MetaDrug™	Metabolism database	www. genego.com
	Metabolite prediction	
	Metabolite prioritization	
	QSAR models for enzymes, transporters and network building algorithms for systems-ADME/Tox	
METEOR™	Rule based metabolite prediction software	www. Ihasalimited.org
	Predicts the metabolite fate of chemicals	
	Displays results as a metabolic tree. User can filter results for likely metabolite	
	Links directly to MetaboLynx for analysis of mass spectroscopy data	
MetabolExpert™	Rule based metabolite prediction software	www. compudrug.com
	Predicts the most common metabolic pathways in animals, plants or photodegradation	
	Results are presented in a metabolic tree format. Graphical Interface for editing and adding rules	
MetaSite™	Site for metabolism prediction for CYP2C9 and CYP3A4 and others	www.moldiscovery.com

4.10 Approaches for computational models for drug metabolism

The approaches used for the computational predictions of human metabolism are

- Databases
- Electronic models
- QSAR/QSMR
- Rule-based approaches
- Machine learning tools
- Pharmacophore modeling
- Homology models

BIOPRESS

4.11 Renal excretion of drugs – Computational models

Renal excretion of drugs can also be predicted by in silico techniques

- Low molecular weight (<500 Daltons) permits the glomerular filtration of drugs
- Dissociation constant decides the drug excretion. All wate soluble drugs or metabolites get excreted in the urine.

4.12 Transporter systems-Modeling-Drug excretion

MRP-1 inhibitors pharmacophore model

All the database molecules used in the modeling contains a ring. This model is used to screen the literature database of clinically useful drugs. In the pharmacophore modeling:

- MRP-1 is inhibited by several compounds like dehydrosilybin, indolopyrimidine and phenoxymethylquinaxalone II
- The pharmacophore model search against the database retrieved candesartan, eprosartan, fexofenadine, losartan, sulfasalazine, telmisartan, vancomycin and zafirlukast as high affinity substrates

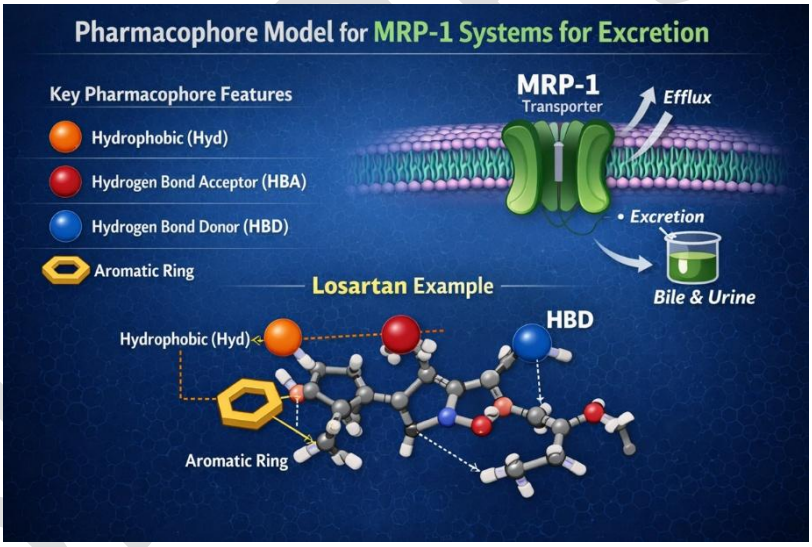


Figure 4.4: Pharmacophore model for MRP-1 system

Organic cation transporters (OCT) influx

The organic cations include drugs and metabolites carrying a net positive charge at physiological pH. OCT facilitate the uptake of many cationic drugs across the barrier membranes form kidney, liver and intestine epithelium.

Organic anion transporter polypeptide (OATP)

These are responsible for actively transporting the anions present in the plasma and thus promoting drug excretion. These transporters also transport organic cationic drugs. The OATP is also present in the tissue membranes of liver, intestine, lung and brain. These transporters mediate sodium dependent transport of a diverse range of amphiphilic organic compounds with

relatively high molecular weight. It includes bile acids, steroid conjugates, thyroid hormones, anionic peptides and numerous drugs.

4.13 Rules and classification systems

Computational methods are those methods in which molecular properties are calculated in order to reflect the behavior of drug molecules.

They are grouped as

- **Substrate based approach**
- QSAR
- Virtual screening
    - Simple counting method (Knowledge based approach)
    - Functional group filters
    - Topological drug classification
    - Pharmacophore point filter
- Machine learning tools
- **Transporter based approach**
- 3D pharmacophore building
- Homology modeling
- De novo approach and Ab initio modeling

Substrate based approach

QSAR

In substrate-based approach, QSAR, QSPR, multivariate analysis and statistical algorithms are used to correlate the molecular descriptors for ADMET related properties. Several QSPR models are developed based on the following strategies

- Membrane interaction-QSAR (MI-QSAR)
- Random forest (RF regression
- Partial least squares (PLS) using 2D descriptors
- Support vector machines (SVMs)
- Artificial neural networks (ANNs): Associative neural networks (ASNNs)
- Bayesian regularized neural networks (BRNNs)

Virtual screening

It is a computational method used to search libraries of small molecules in order to identify those structures, which are most likely to bind to the same drug target.

Virtual screening prioritizes compounds for synthesis/assay. A new ligand may be discovered on the basis of biological structure. The virtual chemical space of small organic molecules is reduced.

Lipinski rule of five for oral absorption

BIOPRESS

Oral absorption of drugs is effective, if they follow the given characteristics

(i) Drugs with molecular weight of less than five hundred (< 500 Da)

(ii) Drugs with hydrophobicity measure (log P) less than five (< 5)

(iii) Drugs with less than five (< 5) hydrogen bond donors

(iv) Drugs with less than ten (< 10) hydrogen bond acceptors

Table 4.3: Rules and Classification System – Knowledge Based Approaches

Oral absorption rules	Classification systems
Lipinski rules of five (RO5)	Biopharmaceutics classification system (BCS)
Veber Rule	Quantitative BCS (QBCS)
Ghost Rule	Biopharmaceutics drug disposition classification system (BDDCS)
Wenlock Rule	Extended Clearance Concept Classification system (ECCS)
Oprea rule of three	
Monika rule	

Veber rule for oral absorption

Veber analysed the molecular properties responsible for predicting the oral bioavailability. The research concluded the importance of molecular flexibility, polar surface area (PSA) and hydrogen bond count in determining the bioavailability

- Rotatable bonds ≤ 10
- Polar surface area (PSA) ≤ 140
- Hydrogen bond count ≤ 12

Ghost rule for oral absorption

Ghost analysed and reported the drug characteristics for better oral absorption. The absorption and permeability of drugs is effective under the following conditions

- Molecular weight between 160 and 480
- Log P between -0.4 and 5.6
- Molar refractivity between 40 and 130
- Number of atoms between 20 and 70

Wenlock rule for oral absorption

The absorption and permeability of drugs is effective under the following conditions

- Molecular weight < 473 Da
- Log P < 5
- Log D < 4.3
- Hydrogen bond donor < 4

- Hydrogen bond acceptors < 7
- Number of rotatable bonds < 11

Oprea rule of three for oral absorption

The drug like molecules should possess the following qualities

- Number of rings ≥ 3
- Number of rigid bonds ≥ 18
- Number of rotatable bonds ≥ 6

Monika rule for oral absorption

- Molecular weight < 500 Da
- Log P between -2 and 5
- HBD <6
- HBA<11
- Number of rotatable bonds <12
- PSA <110

4.14 Biopharmaceutics Classification System

Solubility criterion

A drug is considered highly soluble, when the highest dose is soluble in 250 ml or less of aqueous medium over the pH range 1-8. The ratio of dose/solubility must be less than or equal to 250 ml. A sufficient number of pH conditions should be evaluated to accurately define the pH solubility profile. These conditions are based on the ionization characteristics of the drug.

Permeability criterion

High permeability drugs are generally those having an extent of absorption greater than 90%, in the absence of documented instability. The permeability of a drug can be determined experimentally. It is assessed using any of the following methods used in pharmacokinetics studies.

- In vivo intestinal perfusion studies in humans
- In vivo or in situ intestinal perfusion studies in animals
- In vitro permeation experiments using excised human or animal intestinal tissue
- In vitro permeation experiments across monolayers of cultured human intestinal cells

Applications of BCS

Class 1 (High solubility and high permeability drugs)

Since class I drugs are rapidly absorbed, the bioavailability of the drug is good. Hence, such drugs are suitable for immediate release dosage forms

Class 2 (Low solubility and high permeability drugs)

Since class 2 drugs have low solubility, the formulation should be modified to improve the dissolution rate

Class 3 (High solubility and low permeability drugs)

It is important to modify the dosage form containing class 3 drugs, in order to maximize the time of contact of the drug

Class 4 (Low solubility and low permeability drugs)

These drugs are liable to have poor oral bioavailability. Hence, such drugs should be formulated as prodrugs or alternative route of delivery must be adopted

Flip flop kinetics

The drugs whose absorption half-life is longer than the elimination half life is called as flip flop kinetics. These groups of predominantly class 3 and class 4

Biopharmaceutics drug disposition classification system may be useful in predicting

- Routes of drug elimination
- Efflux and absorptive transporters effects on oral absorption
- Transporter-enzyme interplay yielding clinically significant effects in the intestine, liver, kidney and brain
- Transporter effects on post-absorptive systemic drug levels

Extended Clearance classification system

It is based on in vitro-in vivo extrapolation from hepatic metabolism and transport processes.

It is useful for

- Predicting the predominant clearance mechanism (rate determining step)
- Determining the therapeutic efficacious dose
- Designing compounds with optimized drug disposition characteristics
- Anticipating the potential of victim drugs based on in vitro drug metabolism and transporter data

4.15 Conclusion

Chemoinformatics is known to help in the design and development of safer and effective drugs using computer technologies. These technologies can also be applied in drug disposition and pharmacokinetics. Thus, the processes such as distribution, drug metabolism, drug excretion and drug absorption are predicted using *in silico* methods.

The computer aided product development establishes a relationship between structure and utilization pathways in the body. ADMET prediction and pharmacokinetic modeling have become indispensable components of modern drug discovery and development.

These approaches enable early identification of potential issues related to absorption, distribution, metabolism, excretion, and toxicity, thereby reducing late-stage failures and

associated costs. In silico ADMET tools provide rapid, cost-effective, and reliable predictions that guide the optimization of lead compounds before experimental validation.

Pharmacokinetic modeling further enhances this process by quantitatively describing drug behavior within biological systems, supporting dose selection, therapeutic efficacy, and safety assessment. The integration of computational predictions with experimental data allows for a more comprehensive understanding of drug disposition and variability among patients.

Overall, the combined application of ADMET prediction and pharmacokinetic modeling accelerates the drug development pipeline, improves decision-making, and contributes to the design of safer and more effective therapeutic agents. Continued advancements in computational methods and data-driven approaches are expected to further refine these tools and expand their impact in precision medicine.

4.16 References

1. Davis, C. L., & Michael, W. (2008). *Pharmaceutical analysis*. New Delhi: Wiley India Pvt. Ltd.
2. Swarbrick, J. (2007). *Encyclopedia of pharmaceutical technology* (3rd ed., pp. 701–742). New York: Informa Healthcare.
3. Djuris, J. (2013). *Computer-aided applications in pharmaceutical technology*. Cambridge: Woodhead Publishing.
4. Ekins, S. (2006). *Computer applications in pharmaceutical research and development*. New Jersey: John Wiley & Sons.
5. Subrahmanyam, C. V. S. (2015). *Biopharmaceutics and pharmacokinetics: Concepts and applications* (2nd ed.). Delhi: Vallabh Prakashan.
6. Gupta, R. (2010). *Computers in pharmacy*. New Delhi: Anmol Publications Pvt. Ltd.

BIOPRESS

Chapter 5

Drug Repurposing and Computational Screening Approaches

Arghya Saha Chowdhury[1], Ilias Ahmed[2], Paragmoni Bora[3]

[1]NEF College of Pharmaceutical Education and Research, Nagaon, Assam, 782001

[2, 3] NETES Institute of Pharmaceutical Science, Kamrup, Assam, 781125

Abstract

Drug repurposing (drug repositioning) has emerged as a strategic response to the rising costs, prolonged timelines, and high attrition rates of traditional de novo drug discovery. By identifying new therapeutic indications for existing approved or investigational drugs, repurposing leverages established safety, pharmacokinetic, and toxicological data to accelerate clinical translation while reducing development risk. This review examines the scientific rationale and major classifications of repurposing, including on-target, off-target, reformulation, and combination strategies. It further highlights experimental approaches such as high-throughput and phenotypic screening, alongside computational methodologies including structure-based virtual screening, ligand-based modeling, network pharmacology, transcriptomics-driven signature matching, and knowledge graph analytics. The integration of artificial intelligence (AI) and high-performance computing (HPC) has significantly enhanced systematic target identification, binding affinity prediction, and large-scale hypothesis generation. Landmark case studies across oncology, rare diseases, neurodegeneration, and global health emergencies demonstrate the clinical and economic impact of this approach. Despite regulatory and intellectual property challenges, drug repurposing represents a cost-effective, scientifically robust pathway for accelerating therapeutic innovation in modern precision medicine.

Keywords: Drug repurposing; Drug repositioning; Computational screening; Artificial intelligence; Network pharmacology; Virtual screening; Precision medicine; Drug discovery.

5. Introduction

5.1 Definition and Conceptual Framework of Drug Repurposing

Drug repurposing also termed drug repositioning, reprofiling, or therapeutic switching is the systematic process of identifying new clinical indications for existing drugs that have already been approved by regulatory bodies or have reached advanced stages of clinical evaluation. This strategy capitalizes on the principle of biological pleiotropy, the capacity of a single chemical entity to interact with multiple biological targets or influence diverse pathological pathways.(Al Khzem & Wali, 2025)

Technically, a distinction exists between drug repurposing and API (Active Pharmaceutical Ingredient) redevelopment. While the former often implies using a particular unchanged off-the-shelf drug for a new purpose, medical necessity frequently requires the redevelopment of the API into new dosage forms or release kinetics to better serve the new target population. By late 2025, this field has evolved into programmable medicine, where advanced computational models integrate multi-omics and clinical data to tailor these existing molecules for specific patient subsets.(Ocana et al., 2025)

5.2 Historical Evolution: From Serendipity to Systematic Strategy

Historically, drug repurposing was the result of serendipity or astute clinical observation. Classic examples include Sildenafil, originally developed for angina pectoris but repositioned for erectile dysfunction and later pulmonary arterial hypertension (PAH) after side effects were observed in trials. Similarly, Thalidomide transitioned from a notorious sedative to a standard treatment for multiple myeloma and leprosy complications.(Ghofrani et al., 2006; Hattori & Iguchi, 2004)

In the modern era, the industry has shifted from fortunate accidents to a systematic, data-driven discipline. This transformation is powered by the convergence of high-performance computing (HPC) and artificial intelligence (AI), moving from macroscopic observations to microscopic resolutions of molecular structures. This systematic approach, often called target fishing, allows researchers to mine vast databases to uncover hidden polypharmacology the unrecognized therapeutic potential of established drugs.(Conway & Arlene, n.d.; Keramida et al., 2025)

5.3 Importance in Modern Drug Discovery and Global Health

Repurposing is essential for addressing global health challenges where traditional pipelines are impractical. It is a durable component of innovation a longitudinal analysis of FDA approvals between 1985 and 2024 revealed that 451 drugs received subsequent approval for new therapeutic indications.(Akodad et al., 2026)

Table 5.1: Strategic Impact of Drug Repurposing in Rare Diseases and Global Health Emergencies

Feature	Impact in Rare Diseases	Impact in Global Emergencies
Unmet Need	95% of rare diseases lack approved treatments.	Rapid response required for pandemics like COVID-19.
Success Stories	Lonafarnib repurposed from oncology to Progeria, reducing mortality from 27% to 6%.	Remdesivir (Ebola) and Baricitinib (Arthritis) rapidly deployed for COVID-19.

Economic Logic	Provides a viable R&D path for small patient populations.	Accelerated approval pathways (e.g., 505(b)(2)) bypass redundant trials.

Recent therapeutic shifts illustrate its growing reach. For instance, in 2025, Ozempic (Semaglutide) obtained FDA approval for a therapeutic shift from endocrinology to nephrology to reduce the risk of kidney disease progression in patients with type 2 diabetes.(Aliabadi et al., 2025)

5.4 Traditional Drug Discovery vs. Repurposing: The Attrition Crisis

The primary distinction lies in the risk-reward profile and development timeline. Traditional *de novo* discovery is a lengthy, high-risk process facing an attrition crisis, where approximately 90% of candidates fail during clinical trials.(Ocana et al., 2025)

Table 2: Comparative Metrics of Development Paradigms (2025 Estimates)

Metric	Traditional De Novo Discovery	Drug Repurposing Strategy
Average Cost	$2.0 Billion – $4.5 Billion	~$300 Million
Time to Market	10 – 17 Years	3 – 12 Years (Average 7.2)
Success Probability	~10% approval rate from Phase I	~30% approval rate for de-risked candidates
Early Safety Risk	High; requires full Phase I testing	Minimal; uses established safety data

5.5 The Transformative Role of Computational Tools

Advanced computational tools allow for in silico screening of millions of drug-disease associations that would be impossible to test experimentally. The migration to cloud-based HPC platforms has been a game-changer; contemporary platforms can evaluate molecular databases containing over 100 million compounds within 24–48 hours, a process that previously required several months.

The integration of AI has reduced the time for candidate identification by 71%. Key computational advancements as of 2025 include:

- **Molecular Docking:** Tools like AutoDock Vina now use parallelized, cloud-enabled workflows to process up to 450 million docking operations per day.
- **Protein Structure Prediction:** Breakthroughs like AlphaFold 3 predict joint structures

of complexes including proteins, nucleic acids, and small molecules with unprecedented accuracy.

- **Network Pharmacology:** Modeling biological systems as complex graphs to identify multi-target drugs that can regulate entire nodes in disease-related networks, particularly useful in multifactorial conditions like Alzheimer's.(Ren et al., 2025)

5.6 Rationale for Drug Repurposing

The global shift toward drug repurposing is fundamentally driven by a systemic productivity crisis in traditional pharmaceutical R&D, a phenomenon widely characterized as Eroom's Law. This term, which represents Moore's Law in reverse, observes that the number of new drugs approved per billion dollars of R&D spending has halved approximately every nine years for several decades.(Youngs, 2025)

5.6.1 Economic Imperatives and Development Timelines

The most significant barrier to traditional *de novo* discovery is the prohibitive cost, which is now estimated to exceed $2.5 billion per new chemical entity when accounting for failure costs. In contrast, drug repurposing strategies can bring a therapy to market for approximately $300 million, representing a nearly 85% reduction in capital expenditure. This financial efficiency is largely a function of compressed timelines; while the marathon of *de novo* discovery typically spans 10 to 17 years, repurposing is a comparative sprint requiring only 3 to 12 years. On average, repurposing can slash 5 to 7 years from the standard development clock.(Islam, 2025)

Table 5.3: Economic and Technical Comparison of Development Paradigms

Parameter	Traditional De Novo Development	Drug Repurposing Strategy
Average Cost	$2.0 Billion – $3.0 Billion	~$300 Million
Development Time	10 – 17 Years	3 – 12 Years
Success Probability	~10% approval rate from Phase I	~30% approval rate for de-risked candidates
Safety Threshold	Extensive preclinical & Phase I testing required	Established human safety and PK/PD profiles

Application	Broad, but high risk for niche markets	Global pandemics, rare diseases, oncology

5.6.2 Risk Mitigation and Safety Pre-validation

The traditional drug pipeline faces an attrition crisis where approximately 90% of drug candidates fail during clinical trials, frequently due to unforeseen human toxicity or a simple lack of efficacy. Pipeline attrition rates continue to grow, with the success rate for Phase 1 drugs plummeting to just 6.7% in 2024, compared to 10% a decade ago. Repurposing mitigates these risks because the candidate molecules have already undergone extensive preclinical toxicology and Phase I clinical trials to establish human safety. By starting with de-risked compounds, the probability of regulatory approval triples, rising from 10% to approximately 30%. Success is more likely when there is a clear mechanistic rationale and dose compatibility with established safety profiles.(DrugPatentWatch, 2025)

5.6.3 Regulatory Facilitators: The 505(b)(2) Pathway

Regulatory agencies have responded to these efficiencies by providing streamlined approval pathways. In the United States, the FDA's 505(b)(2) pathway is particularly critical for repurposed products. This hybrid regulatory route allows sponsors to rely on existing scientific literature or the FDA's prior findings of safety and effectiveness for already approved drugs, even if the developer does not own the original data. Between 2017 and 2023, more than half of the 776 chemical NDAs approved by the FDA were granted through the 505(b)(2) pathway, reflecting a clear regulatory preference for strategies that build on existing knowledge rather than duplicating established studies.(Aliabadi et al., 2025)

5.6.4 Urgent Unmet Needs in Rare Diseases and Global Crises

The humanitarian rationale for repurposing is most evident in the treatment of rare and orphan diseases. Approximately 300 million people worldwide live with rare diseases, affecting 3.5% to 5.9% of the global population. Because only 5% of the 7,000 known rare diseases have an approved therapy, the traditional R&D model is often financially unviable due to the small patient volumes. Repurposing allows for the rapid identification of treatments for these populations by leveraging limited datasets from rare disease cohorts. Furthermore, repurposing is a cornerstone of emergency medicine, as seen during the COVID-19 pandemic, where the rapid deployment of agents like remdesivir (originally for Ebola) and dexamethasone demonstrated how existing pharmacopeia can address novel global threats when time is the most critical constraint.(Mucke, 2025)

5.7 Types of Drug Repurposing

Drug repurposing strategies are broadly categorized based on whether the therapeutic effect is achieved through the drug's original target (on-target) or through newly identified molecular interactions (off-target), as well as through technical advancements in formulation and combination.

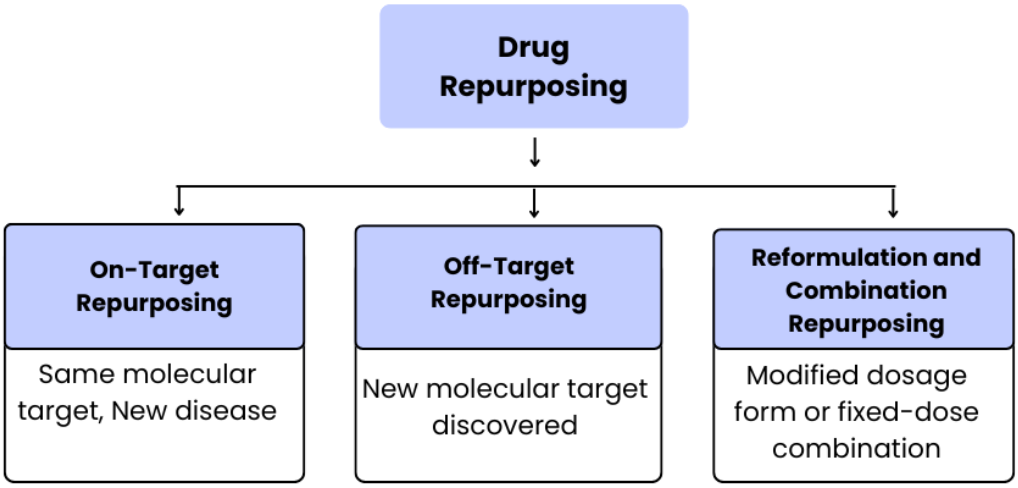


Figure 1. Classification of Drug Repurposing Strategies

5.7.1 On-target Repurposing

On-target repurposing involves leveraging the established pharmacological mechanism of a drug applying its known primary target to a different disease context where that same target plays a pathological role (Eurasian Experiment, 2025). This approach requires ingenuity to connect the dots between varied biological pathways sharing a common molecular node.(Ashburn & Thor, 2004)

- **Aspirin (Acetylsalicylic Acid):** Originally developed as an analgesic and antipyretic, aspirin was later found to inhibit platelet aggregation through the same cyclooxygenase (COX) inhibition pathway. This established its secondary role in the prevention of myocardial infarction and stroke.

- **Finasteride:** Developed initially to treat benign prostatic hyperplasia (BPH) by blocking the 5-alpha-reductase enzyme to limit the bioavailability of dihydrotestosterone (DHT). It was repositioned for male pattern baldness (alopecia) because reducing DHT in the scalp prevents hair follicle miniaturization.

- **Ozempic (Semaglutide):** Initially FDA-approved in 2017 for glycemic control in type 2 diabetes, Ozempic received a transformative repurposing approval in 2025. It is now indicated to reduce the risk of kidney disease progression and cardiovascular death in patients with type 2 diabetes and chronic kidney disease (CKD), shifting its clinical utility from endocrinology to nephrology.

5.7.2 Off-target Repurposing

Off-target repurposing relies on newly discovered mechanisms or those previously considered only in terms of a drug's side effects. This is considered highly innovative and often involves target fishing to uncover unrecognized therapeutic potential within established pharmacopeia.

- **Mebendazole:** Originally an antihelminthic, mebendazole has been identified as a potent anticancer agent. Using the MolTarPred in silico target prediction algorithm, researchers discovered it acts as a potent inhibitor of hMAPK14, a key factor in tumor progression. In vitro validation in glioblastoma cells confirmed binding to the catalytic site of MAPK14 and inhibition of tumor spheroid growth.(Aliabadi et al., 2025; Odarenko et al., 2025)

- **Itraconazole:** A widely used antifungal agent found to exhibit anti-angiogenesis properties. Its ability to modulate pathways independent of its antifungal role led to Phase II trials evaluating it as a second-line therapeutic agent for lung cancer.

- **Ponatinib:** An approved tyrosine kinase inhibitor for leukemia that has been computationally repurposed as a PD-L1 inhibitor. Molecular docking and virtual screening revealed its binding to PD-L1, which was subsequently confirmed to delay tumor growth in mouse models.(Joshua Anthony Vomo et al., 2025)

5.7.3 Drug Reformulation and Combination Repurposing

Advanced repurposing often necessitates modifying the original drug presentation to better serve a new target population or combining it with other agents to achieve synergistic effects.

Table 5.4: Emerging Case Studies in Reformulation and Combination (2024–2025)

Drug / Combination	Mechanism of Innovation	Repurposed Indication
Nicardipine	Bench-to-bedside oral reformulation for pediatrics	Pitt–Hopkins Syndrome (Rare CNS)
Cobenfy™ (Xanomeline + Trospium)	Synergistic agonist/antagonist FDC	Schizophrenia (Approved 2024)
Alyftrek® (Triple modulators)	Strategic deuteration + macrocyclic design	Cystic Fibrosis (Approved 2024)
Donepezil	σ1 agonist property activation	Krabbe Disease (Leukodystrophy)

Brincidofovir	Lipid-conjugate prodrug optimization	Multiple Sclerosis (Patent claimed 2025

- **Nicardipine:** Traditionally used for adult hypertension, nicardipine was repurposed for Pitt-Hopkins syndrome, a rare neurodevelopmental disorder. Computational modeling revealed it could inhibit Nav1.8 neurons, which are hyperexcitable in this condition; oral administration led to measurable developmental improvements in a young patient. (Somorai et al., 2025)

- **Cobenfy™ (Xanomeline-Trospium):** This landmark 2024 approval represents a dual-strategy repurposing. It combines xanomeline (an M1/M4 muscarinic agonist initially developed for Alzheimer's) with trospium (a peripherally restricted muscarinic antagonist). This combination blocks xanomeline's peripheral side effects while preserving its therapeutic antipsychotic effects in the brain.(Kaul et al., 2026)

- **SGLT2 Inhibitors (Canagliflozin):** Originally for diabetes, canagliflozin is being repurposed to treat endometrial cancer by overcoming progestin resistance through the RAR-β/CRABP2 signaling pathway.(Kaul et al., 2026; Yang et al., 2025)

5.8 Approaches to Drug Repurposing

Modern drug repurposing utilizes a synergistic framework that bridges high-throughput experimental validation with advanced computational prediction models.

5.8.1 Experimental Approaches

Experimental methodologies provide the empirical foundation for drug repurposing, offering high-confidence evidence of biological activity in human-relevant models.

High-Throughput Screening (HTS)

High-throughput screening involves the use of automated robotics and sensitive detection tools to rapidly evaluate vast libraries of compounds including FDA-approved drug libraries against specific disease targets or cell lines (Mack, 2024). Modern HTS platforms have transitioned toward miniaturization, using microplates and microfluidic lab-on-a-chip systems to reduce reagent requirements and costs (Singh et al., 2024). A notable application in 2025 was the screening of 2,130 FDA-approved drugs which identified colchicine as a potent therapeutic candidate for atypical teratoid/rhabdoid tumors (AT/RTs), demonstrating IC50 values two orders of magnitude lower than those required for toxicity in normal brain cells.(Kanjanasirirat et al., 2025)

BIOPRESS

Phenotypic Screening

Phenotypic Drug Discovery (PDD) is a mechanism-agnostic approach that identifies bioactive compounds based on their ability to alter an observable characteristic (phenotype) of a cell or organism without prior knowledge of the specific molecular target. This strategy is particularly effective for rare diseases and complex neurological disorders where the underlying pathology is not fully understood. By using patient-derived primary cells or induced pluripotent stem cell (iPSC)-derived organoids, researchers can maintain intact signaling pathways to discover first-in-class repurposing opportunities that might be missed by target-centric methods.(He et al., 2025; Pinzi et al., 2024)

Table 5.5: Key Experimental Models for Drug Repurposing (2025)

Model Type	Primary Application	Advantage in Repurposing
3D Organoids	Oncology and CNS research	Better mimics human tissue architecture and drug penetration.
Organ-on-Chip	Toxicity and PK/PD studies	Recapitulates human physiological processes in a microscale environment.
Zebrafish	Neuroactive drug screening	High genetic similarity to humans; suitable for whole-organism toxicity.
Skin Fibroblasts	Rare metabolic disorders	Convenient access to patient-specific cells for personalized screening.

5.9 Computational Approaches

Computational strategies, collectively referred to as *in silico* repurposing, leverage big data and AI to reduce the search space and mitigate the high attrition rates of traditional discovery.(Islam, 2025)

Systematic Target Fishing

Target fishing uses AI and knowledge graphs to identify potential drug targets based on chemical similarity to known ligands or structural compatibility with protein binding pockets. Breakthroughs in deep representation learning now allow for precise capture rather than broad-spectrum screening, facilitating the discovery of hidden polypharmacology. For instance, the MolTarPred algorithm utilizes Morgan fingerprints and Tanimoto scores to predict interactions

with high accuracy, leading to the identification of hMAPK14 as a target for mebendazole.(Islam, 2025)

In Silico Screening and Molecular Modeling

In silico screening employs high-performance computing (HPC) to simulate drug-target interactions via molecular docking and molecular dynamics (MD). Parallelized pipelines using AutoDock Vina can now perform systematic scans of entire receptor surfaces known as blind docking to detect novel binding sites without prior assumptions. These models integrate physics-based force fields with machine learning to predict binding free energy ($^{\Delta G}$), where more negative values indicate higher stability and therapeutic potential.(Islam, 2025)

Network Pharmacology

Network pharmacology views disease as a perturbation of an intricate molecular network rather than a single-point failure. This holistic approach integrates systems biology, multi-omics data, and graph theory to identify drugs that can modulate multiple nodes simultaneously to restore a diseased network to a healthy state (Mao, 2025). In Alzheimer's research, network-based strategies are used to identify synergistic drug combinations that target amyloid-beta aggregation, tau phosphorylation, and neuroinflammatory pathways concurrently.(Mao, 2025)

Cloud and High-Performance Computing (HPC)

The integration of cloud-enabled HPC has fundamentally transformed development velocities. Modern platforms are capable of processing over 500,000 molecular combinations daily, reducing the preclinical candidate development time to as little as 13–18 months compared to the traditional three to four years. Implementation of cloud-based molecular docking has reduced the computational cost per compound from approximately $9.60 to just $0.34, while maintaining accuracy rates above 98.5%.(Joshua Anthony Vomo et al., 2025)

5.10 Computational Screening Approaches

Modern computational screening for drug repurposing has transitioned from simple pattern matching to high-resolution molecular simulation and autonomous hypothesis generation. These approaches are broadly categorized based on the primary data modality used structure, ligand, network, or transcriptomic profile and are increasingly unified through multimodal foundation models.

Structure-Based Drug Repurposing

Structure-based approaches utilize the three-dimensional (3D) architecture of biological targets to identify potential interactions with existing drugs. This method is increasingly reliant on high-resolution structures from the Protein Data Bank (PDB) or high-accuracy predictions from AlphaFold 3.(Joshua Anthony Vomo et al., 2025)

1. **Preparation:** Retrieve 3D receptor structures from the PDB or generate them via

AlphaFold 3, ensuring proper protonation and charge distribution. Drug libraries are standardized into.pdbqt formats using tools like Open Babel.

2. **Molecular Docking:** Algorithms like AutoDock Vina utilize a gradient-based Iterated Local Search to predict the ideal binding pose of a drug within a target's binding pocket.

3. **Blind Docking:** For targets with unknown active sites, blind docking systematically scans the entire receptor surface to identify novel or cryptic binding pockets.

4. **Affinity Prediction:** The binding free energy ($^{\Delta G}$) is calculated; more negative values indicate higher stability and stronger potential for therapeutic efficacy.

5. **Refinement: Molecular Dynamics (MD)** simulations (e.g., using GROMACS) are conducted for 100 nanoseconds or more to observe the stability of the complex under physiological conditions.(Knox et al., 2024)

A major advancement in 2025 is Nonequilibrium Switching (NES), which offers a 5–10X higher throughput for relative binding free energy (RBFE) calculations compared to traditional equilibrium methods. NES uses many short, bidirectional transitions to provide accurate binding predictions even in systems with trapped water molecules, which often confound standard simulations.(Kapil & Bagga, 2025)

Ligand-Based Drug Repurposing

Ligand-based methods are employed when the 3D structure of a target is unknown or poorly characterized. These methods focus on the chemical and biological properties of known active molecules to search for similar features in approved drug libraries.

- **Pharmacophore Modeling:** This involves identifying the essential 3D arrangement of steric and electronic features (e.g., hydrogen bond donors, hydrophobic regions) required for biological activity.

- **Target Fishing:** Algorithms like MolTarPred use Morgan fingerprints and Tanimoto scores to identify drug targets based on similarity to known ligands. For instance, MolTarPred discovered that the antihelminthic drug mebendazole is a potent inhibitor of hMAPK14, which was later validated as a target for treating glioblastoma.(Odarenko et al., 2025)

- **QSAR Modeling:** Machine learning models are trained on quantitative structure-activity relationship data to predict the biological potency of existing drugs against new therapeutic targets.(Islam, 2025; Pinzi et al., 2024)

Network-Based Approaches

Network pharmacology views disease as a perturbation of a complex molecular web rather than a failure of a single protein. AI frameworks integrate heterogeneous data including drug-disease, drug-target, and protein-protein interaction (PPI) networks to learn high-level representations of drug utility.

- **Random Walk with Restart (RWR):** This algorithm traverses the human interactome network, starting from known disease-associated genes and prioritizing neighbor nodes to find hidden therapeutic links.
- **DREAMwalk:** This model introduces semantic multi-layer guilt-by-association (GBA), allowing a random walker to teleport between semantically related drug and disease nodes based on ATC and MeSH hierarchies.
- **Knowledge Graphs (KGs):** KGs organize biological entities and their relationships as directed edges, facilitating link prediction for novel drug-disease pairings. In 2025, the **TxGNN** model utilized a medical knowledge graph to predict indications across 17,080 distinct diseases, including ultra-rare conditions.(Mao, 2025)

Transcriptomics-Based Repurposing

Transcriptomic approaches rely on the principle of signature matching, hypothesizing that an effective drug will reverse the molecular signature (gene expression profile) of a disease.

- **Connectivity Mapping:** Databases like LINCS L1000 and CMap provide transcriptomic responses for over 42,000 perturbing agents. Researchers use tools like SigCom LINCS to search for reversers drugs that downregulate genes overexpressed in a disease.
- **PETS Framework:** The Predictive Evaluation of Therapeutic Simulations (PETS) is an innovative 2025 in silico framework that integrates pathway models with multi-layer network propagation and memory mechanisms to simulate dynamic drug-induced perturbations over time.(Kapil & Bagga, 2025)
- **scRNA-seq Integration:** Tools like scDrug and scDrugPrio utilize single-cell RNA-sequencing data to predict tumor cell-specific cytotoxicity and prioritize drugs that can overcome resistance to immune checkpoint inhibitors.

Machine Learning and AI Approaches

The 2024–2026 period is defined by the emergence of Foundation Models and Autonomous Scientific Agents that integrate cross-domain biological knowledge.

- **ESM-3:** This flagship multimodal model jointly reasons across protein sequence, structure, and function discrete tokens, enabling the generation of novel protein-binder designs from partial inputs.

- **Reinforcement Learning with Verifiable Rewards (RLVR):** Modern scientific agents use RLVR to explore complex experimental spaces, using code execution or simulation as objective reward signals to refine their strategies.(DrugPatentWatch, 2025)

Table 5.6: Leading AI Platforms and Reasoning Benchmarks (2025–2026)

Platform / Tool	Developer	Capability / Specialty	Performance Benchmark
Omic AI Scientist	Omic.ai	Multimodal omics reasoning	93.3% on GPQA Diamond (PhD-level)
AI Co-Scientist	Google DeepMind	Hypothesis generation agent	Condensed 10-year research into 2 days
Pharma.AI	Insilico Medicine	Generative chemistry & clinical prediction	Phase IIa candidate for IPF in 18 months
Recursion OS	Recursion	Industrialized Phenomics	Screens millions of images for repurposing
Medea Agent	Harvard/MIMS	Tool-aware omics analysis	46% improvement in target identification

These autonomous systems can now simulate thought experiments through text generation, verifying hypotheses against the entire online biomedical literature before proposing a single laboratory test.(Bang et al., 2023)

5.11 Databases and Tools Used in Drug Repurposing

The efficacy of drug repurposing is inextricably linked to the quality, volume, and interoperability of large-scale biomedical and chemical datasets. Modern pipelines integrate diverse data sources from molecular structures and bioactivity data to electronic health records (EHRs) to build predictive models with high clinical transformability.(Bang et al., 2023)

Table 5.7: Key Databases and Computational Tools for Drug Repurposing (2025–2026)

Category	Database / Tool	Core Functionality	Current Data Volume / Impact
Chemical & Bioactivity	**PubChem**	Archive of chemical structures and bioactivity.	123.4 million compounds and 299.4 million bioactivity data points.
	ChEMBL	Manually curated database of bioactive molecules.	17,500 clinical/marketed drugs and 24.2 million activities.
	ZINC	Database of commercially available compounds for virtual screening.	Billions of on-demand drug-like molecules.
Drug-Centric Knowledge	**DrugBank**	Comprehensive pharmacology and drug-target information.	Version 6.0: 4,563 approved and 6,231 investigational drugs.
	DrugCentral	Open-access resource for drug-protein interactions.	Integrated with automated virtual screening pipelines.
Network & Pathway	**STRING**	Global protein-protein association networks.	Version 12.5: 59.3 million proteins across 12,535 organisms.

	KEGG	Biological pathways, diseases, and molecular networks.	Includes 327 human pathway gene sets.
	Cytoscape	Software for network visualization and analysis.	Supports interactive experience for networks with 3M edges.
Structural Biology	**Protein Data Bank (PDB)**	3D structural data for macromolecular complexes.	Over 227,000 structures determined by X-ray and NMR.
	AlphaFold Database	Predicted structures of nearly all catalogued proteins.	Version v6 (2025): 241 million structures including isoforms.
Transcriptomic Signatures	**LINCS L1000**	Dataset of compound-induced gene expression profiles.	1.678 million signatures across 33,621 small molecules.
Clinical & Real-World	**ClinicalTrials.gov**	Database of registered clinical trials and outcomes.	571,979 registered studies as of February 2026.
	OHDSI ATLAS	Real-world data analysis and pharmacovigilance.	Capacity to analyze 100M+ patient records globally.

Integrated Medicinal Chemistry and Bioactivity Repositories

The backbone of computational target fishing is provided by PubChem and ChEMBL. As of 2026, PubChem contains over 123 million unique compounds. Its patent knowledge panel allows users to explore relationships between co-occurring chemicals, genes, and diseases within patent documents. ChEMBL 35 has achieved high accuracy in assay annotation through AI-driven strategies, detecting experimental methods in ~57% of binding assays.(Cuthrell, 2026)

Structural and Network Resources

Structural resources have transitioned from purely experimental data to AI-predicted dark proteome coverage. The AlphaFold Protein Structure Database (AFDB), now synchronized with UniProt 2025_03, provides 241 million high-accuracy predictions. This data is essential for structure-based virtual screening (SBVS) against previously uncharacterized targets. For network analysis, STRING 12.5 offers directionality of regulation, which is critical for understanding information flow in diseased cells.(Szklarczyk et al., 2025)

Real-World Evidence and Clinical Trial Data

The OHDSI (Observational Health Data Sciences and Informatics) initiative has transformed repurposing by enabling the analysis of millions of electronic health records (EHRs). Its ATLAS tool supports the prioritization of repurposing hypotheses through network classification and population-level estimation. Concurrently, ClinicalTrials.gov remains the primary resource for assessing the feasibility and safety history of candidates, hosting over 570,000 interventional and observational studies.(Szklarczyk et al., 2025)

5.12. Case Studies in Drug Repurposing

The successful transition of a drug from its original indication to a repurposed one requires a confluence of mechanistic evidence, established safety data, and often, innovative computational identification. This section examines landmark successes and emerging 2024–2026 candidates.(Sahoo et al., 2021)

Table 5.8: Technical Overview of Successful and Emerging Repurposing Cases

Drug	Original Indication	Repurposed Indication	Primary Mechanism / Status
Sildenafil	Angina Pectoris	ED / PAH / Alzheimer's	PDE-5 Inhibition; neuroprotective (S_R17)

Thalidomide	Morning Sickness	Multiple Myeloma / Leprosy	Immunomodulation / Anti-angiogenic (S_R25)
Remdesivir	Ebola Research	COVID-19	Viral RNA polymerase inhibition (S_R38)
Metformin	Type 2 Diabetes	Oncology / Aging research	AMPK Activation / mTOR inhibition (S_R134)
Lonafarnib	Oncology	Progeria	Farnesyltransferase inhibition (S_R48)
Ozempic	Type 2 Diabetes	Chronic Kidney Disease	GLP-1 receptor agonism (Approved 2025) (S_R269)
Doxycycline	Antibiotic	Alzheimer's Prevention	Fibril disruption (PPI network-identified) (S_S13)
Brincidofovir	Antiviral	Multiple Sclerosis (MS)	Reduction of EBV-infected B cells (S_R21)

Sildenafil: From Cardiovascular to Neurological Applications

Originally identified in 1986 as a pyrazolopyrimidine inhibitor of phosphodiesterase 5 (PDE5) for angina pectoris, sildenafil (UK-92,480) transitioned to erectile dysfunction (ED) in 1998 and pulmonary arterial hypertension (PAH) in 2005. As of early 2026, sildenafil has emerged as a high-priority candidate for Alzheimer's disease prevention. Computational and mouse-model studies indicate it can reduce tau protein buildup and increase cerebral blood flow, with a 2026 Delphi consensus identifying it as a top priority for clinical exploration in cognitive protection.(Alzoubi et al., 2013; Ghofrani et al., 2006)

Thalidomide: Redemption through Mechanism Elucidation

Thalidomide represents one of the most significant drug rescue stories in pharmacological history. After its withdrawal in the 1960s due to severe teratogenicity, it was discovered to possess potent anti-inflammatory and anti-angiogenic properties (Hattori, 2004). This

mechanistic insight led to its repurposing for erythema nodosum leprosum and later multiple myeloma, where it modulates the bone marrow microenvironment and induces tumor cell apoptosis.(Hattori & Iguchi, 2004)

Remdesivir: Rapid Deployment in Global Crisis

Developed initially for Ebola and Marburg virus infections, remdesivir's repurposing for COVID-19 in 2020 demonstrated the power of established safety profiles during a pandemic(Al Khzem & Wali, 2025). As an adenosine analog, it functions by inhibiting the viral RNA-dependent RNA polymerase, thereby suppressing the production of proinflammatory cytokines. This success underscores the synergy between public-private partnerships and repurposing pipelines in addressing emerging infectious threats.(Pauly, 2024)

Metformin: A Pleiotropic Adjuvant in Oncology

Metformin, the standard-of-care for type 2 diabetes, is currently the focus of over 60 active oncology trials. Its anticancer mechanism involves the activation of AMP-activated protein kinase (AMPK) and subsequent inhibition of mTOR signaling, which attenuates insulin-mediated oncogenic cascades. Recent 2025 research further indicates that biomarker-guided strategies involving the PI3K/AKT/mTOR pathway can identify patient subgroups most likely to benefit from metformin as a low-cost cancer therapy.(Yang et al., 2025)

Lonafarnib: First-in-Class Treatment for Progeria

Lonafarnib was initially developed as a cancer therapeutic targeting farnesyltransferase. In 2022, it received landmark approval for Hutchinson-Gilford Progeria Syndrome (HGPS), an ultra-rare fatal premature aging disease. HGPS is caused by the accumulation of a mutant farnesylated protein called progerin; lonafarnib inhibits this post-translational modification, preventing progerin from intercalating into the inner nuclear membrane. Clinical data showed a reduction in mortality from 33.3% to 3.7% after a median of 2.2 years of treatment.(Aliabadi et al., 2025)

Ozempic (Semaglutide): Therapeutic Shift to Nephrology

Semaglutide achieved a transformative repurposing milestone in January 2025 when the FDA expanded its indication from glycemic control in type 2 diabetes to the reduction of kidney disease progression in patients with chronic kidney disease (CKD). Data from the Phase IIIb FLOW trial demonstrated a 24% relative risk reduction in end-stage kidney disease and cardiovascular death, marking a shift in its primary therapeutic utility from endocrinology to nephrology.

7.7 Brincidofovir: Targeting EBV for Multiple Sclerosis

Brincidofovir is a lipid-conjugate prodrug of cidofovir, originally developed for double-stranded DNA viruses. In 2025, patent claims emerged for its use in treating Multiple Sclerosis (MS) based on its ability to significantly reduce the survival of Epstein–Barr virus (EBV)-

infected B cells. Preclinical studies in marmosets have validated its efficacy in limiting neuroinflammation, positioning it as a potential disease-modifying treatment for patients with reactivated EBV.(Corbett et al., 2025)

5.13 Advantages of Drug Repurposing

The shift toward drug repurposing is not merely an incremental improvement in research strategy but a fundamental reorganization of the risk-reward calculus in pharmaceutical development. By leveraging the immense hidden value of already created medicines, the industry can bypass systemic inefficiencies of the *de novo* model.(Mucke, 2025)

Economic Efficiency and Cost Neutralization

The financial logic of repurposing is underscored by its ability to neutralize Eroom's Law, where the number of approved drugs per billion dollars spent has halved every nine years.(Youngs, 2025)

- **Development Cost Neutralization:** While the average fully capitalized cost to bring an entirely new molecular entity (NME) to market is estimated between $2.3 billion and $2.8 billion, a repurposed drug can reach the clinic for approximately $300 million, representing an 85%–88% reduction in capital expenditure.(Al Khzem & Wali, 2025)

- **IT and Infrastructure Savings:** The migration of repurposing pipelines to cloud-based High-Performance Computing (HPC) has reduced computational costs per compound from $9.60 to $0.34 (ResearchGate, 2024). Pharmaceutical organizations leveraging cloud infrastructure have achieved overall IT infrastructure cost reductions of 35%–40%.(Al Khzem & Wali, 2025)

Accelerated Timelines and Rapid Clinical Deployment

The most critical advantage of repurposing is the significant compression of the development cycle, moving from a marathon to a sprint.

- **Cycle Compression:** De novo discovery typically spans 10 to 17 years; in contrast, drug repurposing achieves approval within 3 to 12 years, frequently slashing 5 to 7 years from the typical development clock.(Al Khzem & Wali, 2025)

- **AI-Driven Speed:** AI-enabled workflows demonstrably compress early discovery timelines by 30%–40% (Singh, 2025). High-performance computational frameworks can identify drug candidates for complex diseases like Ebola or Multiple Sclerosis in less than a day.

- **Emergency Responsiveness:** This speed was exemplified during the COVID-19 pandemic, where the rapid deployment of agents like remdesivir (originally for Ebola) and dexamethasone demonstrated how repurposing can provide timely solutions in global health emergencies.(Al Khzem & Wali, 2025)

De-Risking the Pipeline: Safety and Attrition Advantages

Drug repurposing drastically lowers initial R&D risks by leveraging existing data packages.

- **Safety Pre-validation:** Because candidate molecules have already undergone extensive preclinical toxicology and Phase I human safety trials, the risk of toxicological failure is significantly minimized.(Al Khzem & Wali, 2025) This allows researchers to focus primary resources on verifying efficacy for the new indication.

- **Probability of Success:** The approval probability for *de novo* candidates entering Phase I is approximately 10%–11%; for repurposed compounds that have been de-risked, this success rate triples to approximately 30%.(Al Khzem & Wali, 2025)

Table 5.15: Comparative Success and Efficiency Metrics (2025)

Metric	Traditional Development	Drug Repurposing Model
Capital Required	>$2.0 Billion	~$300 Million
Time to Market	10–17 Years	3–12 Years
Phase I Success Rate	40%–65%	80%–90% (AI-native)
Overall Approval Rate	~10%	~30%
Primary Failure Cause	Toxicity/Safety	Efficacy/Business reasons

Regulatory and Patient-Centric Advantages

The 2025 regulatory landscape provides a structured pathway to further incentivize repurposing efforts.

- **The 505(b)(2) Advantage:** This hybrid FDA pathway allows sponsors to rely on existing scientific literature or the FDA's prior findings of safety for approved drugs, reducing the need for redundant nonclinical studies. Over half of recent chemical NDAs have leveraged this pathway.

- **Orphan and Rare Diseases:** Repurposing is often the only feasible choice for the 7,000 known rare diseases where patient populations are too small to commercial justify the $2.5 billion *de novo* expense.(Al Khzem & Wali, 2025)

5.13 Limitations and Challenges

Despite the significant economic and clinical potential of drug repurposing, the field faces a multi-faceted array of scientific, legal, and regulatory barriers that can impede the transition from computational hypothesis to clinical reality.

Intellectual Property (IP) and the Legal Paradox

The most formidable hurdle for commercial developers is the central legal paradox of repurposing: a drug compound that is already in the public domain cannot be patented again for its molecular structure. Consequently, innovators must rely on secondary patents, primarily method-of-use or new formulation claims, which are technically more vulnerable and harder to enforce than original composition-of-matter patents.(Pauly, 2024)

Table 5.10: Comparative Protection and Enforceability of Pharmaceutical Patent Types

Patent Category	Subject Matter	Enforceability Level	Strategic Risk in Repurposing
Composition-of-Matter	The active chemical structure (API).	Very High	Usually expired or nearing expiry for repurposing candidates.
Method-of-Use	Specific therapeutic application/indication.	Moderate to Low	Vulnerable to cross-label prescribing and the doctrine of inherent anticipation.
Formulation/Dosage	Novel delivery systems or release kinetics.	Moderate	Can be bypassed by generic manufacturers using legacy versions of the drug.
Combination	Synergistic fixed-dose drug pairings.	High	Requires demonstration of non-obvious synergy rather than simple additive effects.

A critical legal threat is the doctrine of inherent anticipation, where a patent claim for a new use can be rejected if a prior art reference disclosed a process that implicitly resulted in the same biological effect, even if that effect was not recognized at the time.(Aliabadi et al., 2025; January, n.d.)

Economic Free-Rider Problem and Generic Competition

For drugs that are off-patent, the existence of low-cost generics creates a systemic free-rider problem. Because generic versions share the same active ingredient, physicians may legally prescribe the cheaper legacy brand off-label for the newly discovered indication, even if a dedicated repurposed product has been approved. This phenomenon undercuts the market exclusivity required to justify the $300 million investment in confirmatory clinical trials and regulatory fees(Al Khzem & Wali, 2025). As a result, commercial investment in identifying new uses drops to near zero once generic competition begins.(Boyer & Williams, 2025)

Regulatory Complexity and Marketing Authorization

The lack of standardized, harmonized regulatory pathways for off-patent drugs complicates the recovery of R&D costs.(Garcia-Diaz et al., 2025)While the FDA's 505(b)(2) pathway offers a hybrid route for modified drugs, it still requires a tailored strategic approach to bridge existing safety data with new efficacy requirements. Furthermore, academic and non-profit researchers who lead many repurposing projects often lack the infrastructure and resources to act as a Marketing Authorization Holder (MAH) for commercial approval and post-marketing safety responsibilities.(Pauly, 2024)

Dosing Discrepancies and Pharmacokinetic Barriers

A significant scientific limitation is that a drug's original pharmacokinetic (PK) profile may be incompatible with the needs of the new indication(Al Khzem & Wali, 2025). For example, a drug optimized for metabolic conditions may not achieve therapeutic concentrations in the brain due to the blood-brain barrier (BBB), which remains a persistent limitation for CNS-focused repurposing.(Pauly, 2024)

- **Target Concentration Mismatch:** Modulating a target in a different tissue often requires higher doses, which may escalate systemic toxicity beyond the established safety threshold(Al Khzem & Wali, 2025).
- **Dynamic Response Failures:** Many computational predictions fail to translate clinically because they rely on static representations of biological networks and overlook the multi-scale, dynamic nature of drug-induced perturbations.

5.14 Future Perspectives

The horizon of drug repurposing is being redefined by a transition from human-led computational assistance to fully autonomous discovery ecosystems. As of 2026, the integration of generative biology, agentic workflows, and high-fidelity patient simulations is creating a paradigm shift in how existing pharmacopeia is explored for new therapeutic value.

AI Co-Scientists and Self-Driving Labs

The emergence of AI Co-Scientists represents a leap from predictive models to reasoning agents capable of autonomous hypothesis generation. Leading this transition is the multi-agent framework built on the Gemini 2.0 architecture, which utilizes a generate, debate, and evolve workflow to synthesize novel research proposals.This system employs specialized sub-agents including supervisor, reflection, and evolution units to simulate thought experiments and rank candidates based on novelty and biological plausibility. In validation trials, this framework successfully identified repurposing candidates for acute myeloid leukemia that demonstrated tumor inhibition in vitro at clinically relevant concentrations, condensing years of traditional literature review and brainstorming into just 48 hours.(Ocana et al., 2025)

Complementing these reasoning agents is the industrialization of Self-Driving Labs (SDLs), which close the loop between in silico prediction and physical validation. These facilities, such as the Scientific Discovery Factories launched by Atinary Technologies in 2025, operate 24/7 Design-Make-Test-Analyze-Learn (DMTAL) cycles with minimal human intervention. By integrating Bayesian optimization and active learning, SDLs navigate complex parameter spaces such as optimal drug-excipient ratios for reformulations to achieve convergence on successful candidates up to 100 times faster than manual workflows. The maturation of these labs is expected to reduce the preclinical candidate development timeline from the traditional 3–4 years to just 13–18 months.(Kapil & Bagga, 2025; Penadés et al., 2025)

Digital Twins for Personalized Repurposing

The shift toward precision medicine is facilitated by Digital Twins (DTs), which are virtual replicas of biological systems used to simulate treatment outcomes in real-time. A significant breakthrough in 2025 was the development of the Single-cell Digital Twin (scDT) framework, which models disease heterogeneities at the resolution of individual cell nuclei. By analyzing data from over 1.1 million single nuclei across different stages of Alzheimer's disease progression, researchers constructed cell-type-specific gene regulatory networks to prioritize repurposing candidates.(Corbett et al., 2025) This framework successfully identified 13 FDA-approved drugs including ebselen, resveratrol, and imatinib that could potentially reverse the transcriptomic signatures of neurodegeneration in specific brain cell populations .(Venkatapurapu et al., 2026)

Beyond individual patient modeling, digital twins are increasingly used to create Synthetic Control Arms (SCAs) for clinical trials. SCAs utilize historical trial data and AI-simulated patient profiles to reduce the need for traditional placebo groups, thereby addressing ethical concerns and lowering costs. In oncology, digital twins like the COMPASS model allow for the virtual testing of CAR-T cell therapy trajectories, theoretically optimizing dosing schedules for individual patients before physical administration.As these models evolve into more comprehensive, patient-level representations that integrate longitudinal physiological data, they will enable programmable medicine, where existing drugs are dynamically adjusted based on the digital feedback from the patient's virtual counterpart.(Venkatapurapu et al., 2026)

5.15 Conclusion

The transformation of drug repurposing from an era of serendipity to one of systematic, intelligent drive represents a fundamental reorganization of pharmacological innovation. As of 2026, the convergence of high-performance computing, multi-omics synthesis, and agentic AI workflows has positioned repurposing not merely as a cost-saving tactic, but as a critical strategic imperative for the industry. By leveraging the immense hidden value of already created medicines, researchers have effectively neutralized the crushing economic pressures of Eroom's Law, reducing the typical $2.5 billion *de novo* investment to roughly $300 million and compressing the time-to-market by up to seven years.

The technical landscape is currently defined by a shift from broad-spectrum screening to precise capture, driven by deep learning models that can process over 500,000 molecular combinations daily. The deployment of reasoning agents like the AI Co-Scientist and multimodal foundation models such as ESM-3 marks a transition toward autonomous discovery, where hypotheses are debated, simulated, and refined in *in silico* thought experiments before a single physical assay is conducted. Furthermore, the implementation of Single-cell Digital Twins has allowed for the resolution of inter-patient and intra-tumor heterogeneity, enabling cell-type-specific prioritization that was previously unattainable.

Despite this momentum, significant barriers remain. The legal landscape is increasingly complex, with innovators facing the Catch-22 of proving biological novelty for known structures while navigating the doctrine of inherent anticipation. Moreover, the financial free-rider problem presented by generic competition continues to deter private investment in off-patent drugs, highlighting the urgent need for policy interventions such as pull funding and value-based payments.

Ultimately, the future of drug repurposing lies in its integration into the programmable medicine paradigm a future where virtual replicas of patient systems and synthetic control arms optimize dosing and therapeutic trajectories in real-time. By bridging the gap between existing pharmacological knowledge and emerging global health threats, drug repurposing serves as the most viable solution for providing timely, affordable treatments for rare diseases, oncology, and neurodegeneration, ensuring that therapeutic innovation remains both scientifically robust and socially accessible.

5.16 References

1. Akodad, S., Niu, X., Secades, B., & Stevens, H. (2026). Impact of drug repurposing between 1985 and 2024 on pharmaceutical innovation. *Communications Medicine*. https://doi.org/10.1038/s43856-025-01344-1
2. Al Khzem, A. H., & Wali, S. M. (2025). Drug Repurposing as an Effective Drug Discovery Strategy: A Critical Review. In *Drug Design, Development and Therapy* (Vol. 19, pp. 12019–12034). Dove Medical Press Ltd. https://doi.org/10.2147/DDDT.S576701
3. Aliabadi, A., Moradi, S. Z., Abdian, S., Fakhri, S., & Echeverría, J. (2025). Critical dysregulated signaling pathways in drug resistance: highlighting the repositioning of mebendazole for cancer therapy. In *Frontiers in Pharmacology* (Vol. 16). Frontiers Media SA. https://doi.org/10.3389/fphar.2025.1631419

4. Alzoubi, A., Toba, M., Abe, K., O'Neill, K. D., Rocic, P., Fagan, K. A., McMurtry, I. F., & Oka, M. (2013). Dehydroepiandrosterone restores right ventricular structure and function in rats with severe pulmonary arterial hypertension. *American Journal of Physiology - Heart and Circulatory Physiology*, *304*(12). https://doi.org/10.1152/ajpheart.00746.2012
5. Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs Key Points. In *Nature Reviews Drug Discovery* (Vol. 3). https://doi.org/10.1038/nrd1468
6. Bang, D., Lim, S., Lee, S., & Kim, S. (2023). Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-39301-y
7. Boyer, B., & Williams, M. (2025). *From Concept to Policy - Pull Funding for Repurposed Generic Drugs*.
8. Conway, & Arlene. (n.d.). *Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry and Other Interested Parties DRAFT GUIDANCE*. Retrieved https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances
9. Corbett, A., Sultana, J., Stych, K., Mills, R., Cummings, J. L., Williams, G., Ismail, Z., Soto-Martin, M., Mintzer, J., Gauthier, S., Greig, N. H., Noble, W., Killick, R., Lai, M. K. P., Routledge, C., Walsh, F., Fillit, H., Aarsland, D., Lane, R., … Ballard, C. (2025). Drug repurposing for Alzheimer's disease: a Delphi consensus and stakeholder consultation. In *Alzheimer's Research and Therapy* (Vol. 17, Number 1). BioMed Central Ltd. https://doi.org/10.1186/s13195-025-01895-4
10. Cuthrell, K. M. (2026). AI-guided Drug Repurposing for Neurodegenerative and Neuroinflammatory Diseases. *International Neuropsychiatric Disease Journal*, *23*(1), 1–12. https://doi.org/10.9734/indj/2026/v23i1533
11. DrugPatentWatch, B. (2025). *The AI Revolution in Drug Repurposing: A Comprehensive Pipeline Analysis from Target Identification to Clinical and Commercial Validation Section 1: The Imperative for a Paradigm Shift in Drug Development*. https://www.drugpatentwatch.com/blog/
12. Garcia-Diaz, M., Epstein, D., & Espin, J. (2025). Overcoming barriers to off-patent drug repurposing: a lifecycle-based policy solutions. *Frontiers in Pharmacology*, *16*. https://doi.org/10.3389/fphar.2025.1670845
13. Ghofrani, H. A., Osterloh, I. H., & Grimminger, F. (2006). *Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond*. https://www.nature.com/articles/nrd2030
14. Hattori, Y., & Iguchi, T. (2004). Thalidomide for the treatment of multiple myeloma. In *Congenital anomalies* (Vol. 44, Number 3, pp. 125–136). https://doi.org/10.1111/j.1741-4520.2004.00025.x
15. He, T., Caba, K., & Ballester, P. J. (2025). A precise comparison of molecular target prediction methods. *Digital Discovery*, *4*(9), 2548–2558. https://doi.org/10.1039/d5dd00199d
16. Islam, K. M. T. (2025). Drug Repurposing In Modern Drug Discovery: Role of In Silico Study. *Pharmacon: Jurnal Farmasi Indonesia*, 169–187. https://doi.org/10.23917/pharmacon.v22i2.13698
17. January, D. /. (n.d.). *Patenting New Uses for Existing Drugs: A Legal and Strategic Analysis of Drug Repurposing Section 1: Introduction: The Value and Paradox of Second Acts in Pharmacology*. Retrieved https://www.drugpatentwatch.com/blog/
18. Joshua Anthony Vomo, Papireddypalli Sushma, Bukke Geethanjali, B Dhanush, & Kanala Somasekhar Reddy. (2025). Swiss Dock 2025: Docking approaches comparison between attracting cavities and autodockvina. *World Journal of Biology Pharmacy and Health Sciences*, *22*(2), 120–125. https://doi.org/10.30574/wjbphs.2025.22.2.0452
19. Kanjanasirirat, P., Jearawuttanakul, K., Seemakhan, S., Borwornpinyo, S., Wongtrakoongate, P., Hongeng, S., & Charoensutthivarakul, S. (2025). High-throughput screening of FDA-approved drugs identifies colchicine as a potential therapeutic agent for atypical teratoid/rhabdoid tumors (AT/RTs). *RSC Advances*, *15*(16), 12331–12341. https://doi.org/10.1039/d5ra01341k
20. Kapil, S., & Bagga, P. (2025). In Silico Validation of AI-Assisted Drugs in Healthcare. In *Methods in Molecular Biology* (Vol. 2952, pp. 445–458). Humana Press Inc. https://doi.org/10.1007/978-1-0716-4690-8_24

21. Kaul, I., Claxton, A., Sauder, C., Patel, T., Chaturvedi, S., Brown, D., Zhu, H., Marcus, R., Sawchak, S., & Brannan, S. K. (2026). Long-Term Safety and Efficacy of Xanomeline and Trospium Chloride in Schizophrenia: A 52-Week Open-Label Extension Trial. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.20241186
22. Keramida, P., Syrigos, N. K., Kouvela, M., Poulakou, G., Charpidou, A., & Fiste, O. (2025). AI-Driven Drug Repurposing: Applications and Challenges. *Medicines*, *12*(4), 28. https://doi.org/10.3390/medicines12040028
23. Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., Pon, A., Cox, J., Chin, N. E. L., Strawbridge, S. A., Garcia-Patino, M., Kruger, R., Sivakumaran, A., Sanford, S., Doshi, R., Khetarpal, N., Fatokun, O., Doucet, D., Zubkowski, A., … Wishart, D. S. (2024). DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research*, *52*(D1), D1265–D1275. https://doi.org/10.1093/nar/gkad976
24. Mao, J. (2025). Editorial: The role of genes and network pharmacology in new drug discovery. In *Frontiers in Genetics* (Vol. 16). Frontiers Media SA. https://doi.org/10.3389/fgene.2025.1578094
25. Mucke, H. A. M. (2025). Drug Repurposing Patent Applications April–June 2025. *Drug Repurposing*, *2*(2). https://doi.org/10.58647/drugrepo.25.1.0008
26. Ocana, A., Pandiella, A., Privat, C., Bravo, I., Luengo-Oroz, M., Amir, E., & Gyorffy, B. (2025). Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. In *Biomarker Research* (Vol. 13, Number 1). BioMed Central Ltd. https://doi.org/10.1186/s40364-025-00758-2
27. Odarenko, K. V., Zenkova, M. A., & Markov, A. V. (2025). Transcriptomic-Driven Drug Repurposing Reveals SP600125 as a Promising Drug Candidate for the Treatment of Glial-Mesenchymal Transition in Glioblastoma. *International Journal of Molecular Sciences*, *26*(19). https://doi.org/10.3390/ijms26199772
28. Pauly, B. A. (2024). Accelerating Drug Development: Unleashing the Power of Drug Repurposing. *Drug Repurposing*, *1*(1). https://doi.org/10.58647/drugrepo.24.1.0013
29. Penadés, J. R., Gottweis, J., He, L., Patkowski, J. B., Shurick, A., Weng, W.-H., Tu, T., Palepu, A., Myaskovsky, A., Pawlosky, A., Natarajan, V., Karthikesalingam, A., & Costa, T. R. D. (2025). *AI mirrors experimental science to uncover a novel mechanism of gene transfer crucial to bacterial evolution*. https://doi.org/10.1101/2025.02.19.639094
30. Pinzi, L., Bisi, N., & Rastelli, G. (2024). How drug repurposing can advance drug discovery: challenges and opportunities. In *Frontiers in Drug Discovery* (Vol. 4). Frontiers Media SA. https://doi.org/10.3389/fddsv.2024.1460100
31. Ray, R., Shende[2], V., & Sukul[3], S. (2025). *European Journal of Cardiovascular Medicine (EJCM) Association between inflammatory markers (CRP, ESR) and MRI findings in rheumatoid arthritis*. https://www.healthcare-bulletin.co.uk/
32. Ren, Y., Hu, M., Li, Y. E., Pieper, A. A., Cummings, J., & Cheng, F. (2025). *Single-cell digital twins identify drug targets and repurposable medicine in Alzheimer's disease*. https://doi.org/10.1101/2025.01.14.633081
33. Sahoo, B. M., Ravi Kumar, B. V. V., Sruti, J., Mahapatra, M. K., Banik, B. K., & Borah, P. (2021). Drug Repurposing Strategy (DRS): Emerging Approach to Identify Potential Therapeutics for Treatment of Novel Coronavirus Infection. In *Frontiers in Molecular Biosciences* (Vol. 8). Frontiers Media S.A. https://doi.org/10.3389/fmolb.2021.628144
34. Somorai, M. A., Ekins, S., Rupprecht, C., Lettl, C., & Mall, V. (2025). Repurposing nicardipine leads to improved development in a young patient with Pitt–Hopkins syndrome. *Frontiers in Pharmacology*, *16*. https://doi.org/10.3389/fphar.2025.1592011
35. Szklarczyk, D., Nastou, K., Koutrouli, M., Kirsch, R., Mehryary, F., Hachilif, R., Hu, D., Peluso, M. E., Huang, Q., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & Von Mering, C. (2025). The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Research*, *53*(D1), D730–D737. https://doi.org/10.1093/nar/gkae1113
36. Venkatapurapu, S. P., Clegg, L., Nowojewski, A., Kimko, H., Olabode, D., Sawant-Basak, A., & Vishwanathan, K. (2026). Digital twins for accelerating drug discovery and development: opportunities

and challenges. In *Drug Discovery Today* (Vol. 31, Number 2). Elsevier Ltd. https://doi.org/10.1016/j.drudis.2026.104617

37. Yang, Y., Zhou, J., Lv, Q., Ni, Q., Hu, B., Wang, Y., Qu, S., Li, G., Yang, W., Zhong, R., Chen, X., & Zhu, Y. (2025). Canagliflozin alleviates progestin resistance by suppressing RARβ/CRABP2 signaling in THRB knockout endometrial cancer cells. *Frontiers in Pharmacology*, *16*. https://doi.org/10.3389/fphar.2025.1573032
38. Youngs, R. (2025). *Reversing Eroom's Law: Finitude and Dilution in Small-Molecule Discovery*. https://doi.org/10.1101/2025.10.29.685434

BIOPRESS

Chapter 6

Biomarkers in Precision Therapeutics

Dr C. Jaganmohan[1*], Dr Srinidhi. S[2]

1 Professor & Head, Department of Pharmacy Practice. PPG College of Pharmacy, Coimbatore.*

2 Assistant Professor, Department of Pharmacy Practice. PPG College of Pharmacy, Coimbatore.

Abstract

The transition from a standardized, empirical medical model toward a highly individualized paradigm, known as precision medicine is fundamentally underpinned by the discovery, validation, and clinical integration of biomarkers. Biomarkers are defined as objective, measurable indicators of normal biological processes, pathogenic states, or pharmacological responses to therapeutic interventions. Within the context of precision therapeutics, these indicators serve as essential tools for patient stratification, enabling the identification of specific cohorts most likely to respond to a particular drug while minimizing the risk of adverse reactions. This chapter provides an exhaustive examination of the biomarker landscape, beginning with the historical evolution of personalized medicine and the harmonization of terminology established by regulatory bodies such as the FDA and NIH through the BEST (Biomarkers, EndpointS, and other Tools) resource.

The classification of biomarkers is explored through both functional lenses diagnostic, prognostic, predictive, and safety and molecular sources, encompassing the multi-omics triad of genomics, transcriptomics, and proteomics, as well as the emerging field of metabolomics. Technological advancements, particularly high-throughput sequencing and mass spectrometry, have facilitated a systems-level understanding of disease, now further enhanced by the application of artificial intelligence and machine learning to decode high-dimensional datasets. Clinical breakthroughs in oncology, including liquid biopsies and the detection of molecular residual disease, are in contrast with advancements in cardiovascular disease management and the recent surge in blood-based biomarkers for neurodegenerative disorders presented at the 2025 Clinical Trials on Alzheimer's disease (CTAD) conference. Finally, this report addresses the regulatory qualification pathways, the ethical implications of genetic permanence, and the future necessity of extending precision medicine to low- and middle-income countries (LMICs) to bridge the global diagnostic divide.

Keywords

Biomarkers, Precision therapeutics, Personalized medicine, Companion diagnostics, Pharmacogenomics, Omics technologies, Artificial intelligence, Liquid biopsy, Regulatory qualification.

6.1 Introduction

The paradigm shift from conventional "one-size-fits-all" medicine to precision medicine represents one of the most profound transformations in the history of healthcare. While the practice of adapting treatment to an individual's health status was documented as early as 1550 BC in ancient Egyptian medicine and alluded to in the Homeric epics, it remained largely empirical and anecdotal for millennia. Hippocrates' early insights into human distinctiveness and the influence of the "four humors" on disease prediction laid a conceptual foundation that has only recently been translated into a molecular reality[1,2]. The birth of the scientific discipline of pharmacogenetics in the early 1950s provided the first rigorous evidence that drug response variability is largely genetically determined. This field has since evolved into the broader domain of pharmacogenomics, which utilizes the entire molecular landscape of the patient to inform therapeutic decisions[3].

The modern definition of a biomarker, as harmonized by the FDA-NIH Biomarker Working Group, characterizes it as "a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions". It is important to distinguish biomarkers from clinical outcome assessments; while a biomarker is a measurement of a biological state, it is not a direct assessment of how a patient feels, functions, or survives. This distinction is critical in drug development, where biomarkers are often used as surrogate endpoints measures that are reasonably likely to predict a clinical benefit[4, 5]

In the current landscape of drug development, the integration of biomarkers is no longer optional but essential. The completion of the Human genome project (HPG) in 2003 accelerated the pace of discovery, allowing clinicians to move beyond symptom-based classification toward molecular profiling. This approach allows for the identification of disease susceptibility

before symptoms manifest, the selection of optimal therapies to reduce trial-and-error prescribing, and the revival of drugs that may have failed in general populations but show high efficacy in specific genomic subgroups. The objective of this report is to delineate the multi-faceted roles of biomarkers in enhancing the precision of therapeutic interventions and to explore the technological and regulatory infrastructures required to sustain this momentum.[5,6]

6.2 Types and classification of biomarkers

The utility of a biomarker is contingent upon its rigorous classification, which provides the framework for its "context of use" (COU) in clinical and research settings.[7] Regulatory agencies emphasize that the same biological analyte can serve multiple roles depending on the clinical question being addressed.[8]

6.2.1 Based on function

Functional classification focuses on the clinical application and the predictive value of the marker regarding disease state or therapeutic impact.

- **Diagnostic biomarkers:** These are utilized to detect or confirm the presence of a disease or to identify individuals within a specific disease subtype. In neuro-oncology,

the presence of 1DH1 or 1DH2 mutations serves as a primary diagnostic marker to categorize gliomas, which significantly influences subsequent treatment paths.

- **Prognostic biomarkers:** These indicate the likelihood of a clinical event, such as disease recurrence or progression, regardless of the treatment intervention. For example, 1p/19 co-deletion in certain brain tumors is a strong indicator of a more favorable natural history and longer survival.
- **Predictive biomarkers:** These are perhaps the most vital in precision therapeutics, as they identify individuals who are more likely than similar individuals without the marker to experience a favorable or unfavorable effect from a specific medical product. MGMT promoter methylation status is a classic predictive biomarker in glioblastoma, identifying patients who will likely benefit from temozolomide chemotherapy.
- **Pharmacodynamic (Response) biomarkers:** These provide evidence that a biological response has occurred in a patient exposed to a drug. They are used to track whether a drug is hitting its intended target, such as a reduction in malignant cell counts in cerebrospinal fluid following treatment.
- **Safety biomarkers:** These are measured to indicate the likelihood, presence, or extent of toxicity as an adverse effect of a medical product. Monitoring complete blood counts to assess bone marrow suppression or identifying genetic polymorphisms like NQO: alleles to predict drug toxicity are standard safety protocols.
- **Susceptibility/Risk biomarkers:** These identify individuals at risk for developing a condition before they are symptomatic. Genetic testing for BRC A1/2 mutations provides a susceptibility assessment for breast and ovarian cancers, allowing for proactive monitoring or preventive interventions[9,10.]

Table 6.1: Types of biomarkers and their functional objective[9]

Biomarker type	Functional objective	Clinical example
Diagnostic	Disease identification/subtyping	1DH/ mutations in gliomas
Prognostic	Prediction of natural disease course	1p/19 codeletion in oligodendroglioma
Predictive	Predicting response to specific therapy	MGMT methylation for temozolomide

Safety	Monitoring for adverse effects/toxicity	NQO: polymorphisms (rs1800566)
Monitoring	Tracking disease status serially	ctDNA levels for tumor burden

6.2.2 Based on source

Biomarkers can also be classified by the molecular level from which they originate, reflecting the "central dogma" of biology and the layers of complexity within an organism.

- **Genomic biomarkers:** These represent DNA-based characteristics, including single-nucleotide polymorphisms (SNPs), copy number variations, and structural rearrangements. They provide the stable, inherited blueprint of an individual's genetic susceptibility.
- **Transcriptomic biomarkers:** These involve RNA expression patterns, which offer a dynamic view of gene activity in response to stimuli. Transcriptomics captures the "spatio-temporal" gene expression dynamics that DNA alone cannot reflect.[11, 12]
- **Proteomic biomarkers:** These study the entire set of proteins (the proteome). Since proteins are the actual functional effectors and metabolic catalysts, proteomics directly reflects the current biological activity and signaling status of a cell.
- **Metabolomic biomarkers:** These are small-molecule metabolites that serve as the final downstream products of cellular processes. Metabolomics provides a "chemical fingerprint" of metabolic activity, often reflecting the ultimate phenotype of a system.[13]

6.3 Role of biomarkers in precision therapeutics

The integration of biomarkers into clinical practice allows for a transition from population-level averages to individual precision management. By providing tailored risk assessments and intervention recommendations, biomarkers maximize clinical benefit while minimizing exposure to ineffective or harmful treatments.[14]

6.3.1 Patient stratification

The identification of "responders" versus "non-responders" is the cornerstone of clinical trial enrichment. Historically, many drugs failed in Phase III trials because the study population was too heterogeneous. Biomarker-driven stratification ensures that the right patients are enrolled, which not only increases the statistical power of the trial but also reduces the number of participants exposed to unnecessary risk.[15,16] This is particularly evident in oncology, where patients are now routinely stratified based on genomic profiling (e.g., EGFR, ALK or ROS1 status) rather than just the anatomical location of the tumor.[17,18,19,20]

6.3.2 Drug selection

Targeted therapy selection is the primary application of predictive biomarkers. In precision oncology, imatinib's success in treating Chronic myeloid leukemia (CML) harboring the BCR-

ABL rearrangement served as a "poster child" for the field. The selection of tyrosine kinase inhibitors (TKIs) or immune checkpoint inhibitors (ICIs) is now dependent on molecular profiling obtained from tissues or liquid biopsies, ensuring that the therapeutic mechanism aligns with the tumor's driving mutation.[20]

6.3.3 Dose optimization

Pharmacogenomic biomarkers are instrumental in individualized dosing strategies. Genetic variations in metabolic enzymes, such as the cytochrome P-450 (CYP) family, determine the rate at which a patient processes a drug. For instance, a patient identified as a CYP2C9 poor metabolizer may require a significantly lower dose of certain anticoagulants to avoid life-threatening bleeding. This precision minimizes adverse drug reactions and maintains the drug within the narrow therapeutic window.

6.3.4 Monitoring treatment response

The ability to track efficacy in real-time allows for dynamic adjustments to the therapeutic regimen. Monitoring biomarkers, such as serial assessments of circulating tumor DNA (ctDNA) or Brain natriuretic peptide (BNP) in heart failure, provide a non-invasive window into the patient's response.[20] In oncology, the emergence of liquid biopsy allows for the detection of "minimal residual disease" (MRD) long before traditional imaging can detect a recurrence, enabling earlier and more effective intervention.[21]

6.4 Biomarkers in drug discovery and development

Biomarkers have become indispensable tools across all phases of the drug development lifecycle, from target identification to post-market surveillance.

- **Target identification and validation:** High-throughput omics screening allows researchers to identify molecular drivers of disease. Overlapping signals across different layers (e.g., a genetic mutation that correlates with an overexpressed protein) increase confidence in the biological relevance of a potential drug target.[22,23]
- **Clinical trial design:** Using biomarkers as inclusion criteria ensures that the trial population is enriched with patients most likely to respond to the intervention. This "enrichment" strategy significantly reduces the time and cost of clinical trials.[24.25]
- **Companion diagnostics (CDx):** In many cases, the FDA requires a companion diagnostic test to be develop and approved alongside a new drug. This ensures that the drug is only prescribed to patients who have been confirmed to possess the relevant biomarker, thereby safeguarding both efficacy and safety.[26]
- **Regulatory importance:** A qualified biomarker—one that has undergone formal regulatory review for a specific context of use—reduces uncertainty in the approval process. The 21st Century cures act has established a clear, three-stage process for this qualification in the United States, facilitating the use of novel biomarkers as drug development tools.[27]

6.5 Technologies for biomarker discovery

A suite of high throughput technologies that provide a system-level view of biological complexity drives the transition of biomarkers from the bench to the clinic.[28]

6.5.1 Omics technologies

Modern discovery relies on the integration of multiple "omics" disciplines to overcome the limitations of single-analyte studies.

- **Genomics:** Next-generation sequencing (NGS) has become the gold standard for identifying targetable mutations and structural variations. Landmark projects like The Cancer Genome Atlas (TCGA) have demonstrated the utility of genomics in uncovering clinically actionable alterations across various cancer types.[29,30,31,32]
- **Proteomics:** Since proteins are the direct functional effectors of cellular activity, proteomics provides insights into signalling pathways and protein-protein interactions. Technologies like mass spectrometry and aptamer-based SOMA scan allow for the quantification of thousands of protein fragments, revealing metabolic mechanisms that DNA sequencing might miss.
- **Metabolomics:** This field captures the ultimate chemical fingerprints of cellular processes. Metabolomic markers are particularly sensitive to environmental changes, making them valuable for studying how a patient's external environment (e.g., diet, pollutants) interacts with their internal biology.[33]

6.5.2 High-throughput screening (HTS)

HTS technologies allow for the rapid profiling of thousands of compounds or genetic variants in vitro. By treating cell lines or patient-derived organoids with investigational agents and analyzing the resulting omics profiles, researchers can identify gene signatures that correlate with drug response or resistance.[34,35]

6.5.3 Bioinformatics and data analysis

The integration of massive datasets requires advanced computational frameworks. Bioinformatics tools are used to normalize data, correct for batch effects, and map identifiers across different platforms (e.g., mapping a transcriptomic probe to a specific protein). Systems biology approaches then convert these data into interactive networks, identifying regulatory "hubs" that can serve as more robust biomarkers than individual molecules.[36]

6.6 Role of artificial intelligence (AI) and machine learning

The advent of multi-omics has created a "data deluge" that exceeds the capacity of conventional statistical methods. Artificial Intelligence (AI) and Machine Learning (ML) are now essential for deriving clinically actionable insights from these high-dimensional datasets.

6.6.1 Biomarker identification using AI

AI algorithms excel at feature selection—the process of identifying the most predictive markers within thousands of candidates. Embedded methods such as LASSO (Least Absolute Shrinkage

and Selection Operator) are used to integrate feature selection directly into the model training process, effectively reducing noise and preventing overfitting.[37,38]

6.6.2 Predictive modeling

Supervised learning techniques, such as support vector machines (SVM) and random forests, are utilized to classify disease states and predict patient outcomes. SVMs are particularly effective for high-dimensional omics data with small sample sizes, while random forests provide robustness against noise. More recently, Deep learning (DL) models, including

convolutional neural networks (CNNs), are being used to integrate molecular data with medical imaging, providing a truly multimodal view of the patient.[39]

6.6.3 Integration of multi-omics datasets

Integrating disparate layers of biological information—genomics, transcriptomics, and clinical records—requires "intelligent systems" that can learn the latent relationships between these data types. AI-driven pipelines, such as those used in sepsis research or precision oncology, can harmonize heterogeneous data into unified predictive frameworks, narrowing the gap between molecular discovery and clinical application.[39]

Table 6.2: AI/ML model and their primary application in biomarker research[39,40]

AI/ML model	**Primary application in biomarker research**	**Key advantage**
Support vector machines (SVM)	Classification of disease subtypes	Effective for small sample sizes in high-dimensional data
Random forests	Predicting therapeutic response	Robust against data noise and overfitting
LASSO regression	Feature selection from omics datasets	Simplifies models by penalizing less relevant features
Deep learning (CNN/RNN)	Imaging-omics fusion & time-series analysis	Captures complex, non-linear biological relationships
Transformer/LLM	Target discovery from scientific literature	Processes unstructured data to surface novel associations

6.7 Clinical applications of biomarkers

The real-world impact of biomarkers is most evident in the fields of oncology, cardiology, neurology, and infectious diseases, where they have reshaped the standard of care.

6.7.1 Oncology

Oncology remains the most advanced area of precision medicine. The concept of "liquid biopsy" has emerged as a minimally invasive method to profile tumor dynamics using blood samples.

- **Circulating tumor DNA (ctDNA):** ctDNA allows for the detection of mutations and methylation patterns. In a case study (NCT04820868), a multicancer detection model using over 160,000 CpG methylation sites accurately identified and localized six different cancer types.
- **Circulating tumor cells (CTCs):** CTCs provide a clonal relationship to the primary tumor. In prostate cancer trials (NCT01809691), CTC counts were validated as a prognostic biomarker for estimating survival outcomes.
- **Targeted therapies:** The development of imatinib for BCR-ABL1 positive leukemia and TKIs for EGFR mutated lung cancer remains the benchmark for biomarker-driven drug discovery.[41]

6.7.2 Cardiovascular diseases

Cardiovascular medicine is shifting from reactive treatment to proactive risk stratification.

- **Acute coronary syndromes (ACS):** Beyond standard Troponin levels, precision medicine now incorporates machine learning-based diagnostic scores like CoDE-ACS to distinguish between different mechanisms of cardiac injury.
- **PCSK9 targeting:** The management of LDL cholesterol has evolved with the use of monoclonal antibodies (mAbs) like alirocumab and siRNA agents like inclisiran. While both lower LDL-C, their different pharmacokinetic profiles-mAbs working outside the cell and siRNA inhibiting synthesis inside the hepatocyte-offer distinct opportunities for acute vs chronic care.
- **HFpEF phenotyping:** Researchers are using AI to identify "phenotype clusters" in heart failure patients with preserved ejection fraction, allowing for better management of the complex co-morbidities often associated with the condition[39,42]

6.7.3 Neurological Disorders

Breakthroughs in blood-based biomarkers for Alzheimer's disease (AD) were the highlight of the 2025 Clinical trials on Alzheimer's Disease (CTAD) conference.

- **p-tau217:** This marker has emerged as a highly reliable indicator of AD pathology. Studies have now validated capillary blood collection (via finger-prick) for p-tau217 measurement, enabling remote sampling and more frequent monitoring in clinical trials.
- **eMTBR-tau243:** While p-tau markers track amyloid pathology, eMTBR-tau243 is more specifically associated with tau tangle pathology, providing a closer correlation with cognitive decline.
- **Neurofilament light chain (NfL):** NfL has been established as a robust indicator of general neurodegeneration, used to monitor disease modification in conditions like Pompe disease and multiple sclerosis.[38]

6.7.4 Infectious diseases

BIOPRESS

Biomarkers in infectious diseases are vital for differentiating pathogens and predicting severe outcomes like sepsis.

- **Sepsis:** Procalcitonin (PCT) and its ratio to C-reactive protein (CRP) are used to guide antibiotic therapy. Newer genetic markers, such as the long noncoding RNA NEAT1, have been identified as potential early diagnostic indicators of sepsis-related immune dysfunction.[40]
- **COVID-19:** Severe infection is characterized by elevated D-dimer, lymphopenia, and high levels of cytokines like IL-6 and TNF-α. Cardiac biomarkers, including high-sensitivity troponin I (h_8-cT_nI), have also been correlated with thoracic CT findings and adverse outcomes.[41]
- **Tuberculosis (TB):** Host gene signatures (e.g., CD64, GBP5, GZMA) are being explored as blood-based tests that can reliably distinguish active TB from latent infection, which is a major challenge for traditional immunological assays.[43]

6.8 Experimental and validation approaches

Before a biomarker can be integrated into a regulatory submission, it must undergo extensive validation to ensure its accuracy and clinical relevance.

6.8.1 In Vitro and In Vivo Validation

The discovery process often begins with in vitro screening of cell lines or organoids to establish a baseline signature of drug response. This is followed by in vivo validation, often using mouse clinical trials (MCT) with patient-derived xenograft (PDX) models. These studies correlate pharmacological responses with genomic information to confirm the biomarker's predictive value.

The V3 framework for digital biomarkers

For digital biomarkers and wearables, a specialized "V3 framework" is applied:

- **Verification:** Ensures the sensor captures the raw signal correctly (e.g., bench testing an accelerometer).
- **Analytical validation:** Confirms the algorithm transforms raw data into a reliable measure (e.g., comparing algorithm-derived step counts to a manual tally).
- **Clinical Validation:** Determines if the measure answers a specific clinical question in its context of use.[44]

6.8.2 Assay development

Developing a robust assay requires standardized protocols for sample collection, handling, and storage. The stability of the specimen is paramount, as pre-analytical factors can significantly skew results. For clinical implementation, assays must be performed in licensed labs (e.g., CLIA-certified in the US) that adhere to stringent quality control standards.

6.9 Challenges and limitations

Despite the rapid pace of discovery, several hurdles prevent most candidate biomarkers from reaching clinical practice.

- **Variability and reproducibility:** Many biomarkers discovered in small, controlled cohorts fail to replicate in larger, more diverse populations. "Batch effects" in omics data and metadata inconsistencies across different institutions often lead to models that show high internal accuracy but poor generalizability.
- **High cost and complexity:** The infrastructure required for multi-omics analysis—high-resolution mass spectrometers, NGS platforms, and high-performance computing—is expensive to maintain. This limits access primarily to affluent countries.
- **Regulatory challenges:** Qualifying a biomarker for a specific context of use is an arduous process that requires robust evidence from multiple trials. Surrogate endpoints, which predict clinical benefit without being direct measures of it, require the highest level of evidence.
- **Data interpretation:** The "P > N" problem (where the number of predictors P is much larger than the number of subjects N) makes data interpretation complex. Distinguishing "signal" from "noise" in high-dimensional omics data requires sophisticated bioinformatics that may introduce its own biases.[45,46,47]

6.10 Ethical and regulatory considerations

6.10.1 Patient data privacy

Genetic information is unique, permanent, and predictive, raising significant privacy concerns. While the genetic information nondiscrimination act (GINA) prevents insurers and employers from using genetic data for coverage or hiring decisions, it does not cover all types of insurance (e.g., life insurance). HIPAA protects medical records but often does not apply to direct-to-consumer (DTC) testing companies.

6.10.2 Ethical concerns in genetic testing

Disclosure of "incidental findings"-unsolicited discoveries unrelated to the primary study is a major ethical dilemma. Guidelines generally recommend that participants be

informed of this possibility and given the choice to receive such information. Furthermore, the permanence of genetic data means that a breach of confidentiality can have lifelong implications for both the individual and their relatives. [49,50]

6.10.3 Regulatory frameworks

The FDA and EMA have established formal pathways for biomarker qualification.

- **ICH E15 and E16:** These international guidelines harmonize the terminology for genomic biomarkers and define the structure and format for qualification submissions.

- **Qualification advice/opinion:** The EMA offers a voluntary pathway for the qualification of novel methodologies in drug development, facilitating early engagement with developers to speed up the process. [51,52]

6.11 Emerging trends

6.11.1 Liquid biopsy and MRD

Liquid biopsies are increasingly being used not just for diagnosis but for the longitudinal monitoring of cancer. The detection of molecular residual disease (MRD) after surgery allows for a more personalized approach to adjuvant therapy, identifying patients who may benefit from further treatment vs. those who can safely avoid it.

6.11.2 Digital biomarkers and wearables

Digital biomarkers derived from sensors in smartphones and wearables provide continuous, real-world monitoring of physiological and behavioral metrics. Examples include fall-risk alerts from gait analysis in Parkinson's disease or the detection of atrial fibrillation from wearable ECGs.

6.11.3 Real-time monitoring systems

The integration of biochemical sensing with digital technology, such as continuous glucose monitoring (CGM), allows for data-driven, real-time insulin dosing. This synergy between "wet" biomarkers and digital processing is expected to expand into other areas of chronic disease management.[53]

6.12 Future perspectives

6.12.1 Fully personalized therapeutics

The future of precision medicine involves "digital twins"-real-time simulations of a patient’s physiological data to optimize intervention strategies before they are applied. This will allow for the optimization of dynamic therapies that adapt to the patient's changing state.

6.12.2 Integration of AI and clinical practice

AI will move from being a research tool to a real-time clinical assistant. "Explainable AI" mechanisms will be vital for clinical translation, providing doctors with interpretable rules for why a certain treatment was recommended based on the integrated multi-omics profile of the patient.

6.12.3 Precision medicine in developing countries

The biggest determinant of cancer outcome currently is one’s place of birth. Bridging the "diagnostic divide" will require heavy investment in pathology services and the adoption of simplified, equipment-free diagnostics (like LAMP) that can operate in low-resource settings. Partnerships like those between The Max Foundation and Novartis for CML treatment demonstrate that precision medicine is possible in developing nations through creative solutions and global collaboration.[53,54,55]

BIOPRESS

6.13 Conclusion

Biomarkers have evolved from simple laboratory measurements to complex, multi-dimensional indicators that drive the entire cycle of modern medicine. By enabling the precise stratification of patients, the selection of targeted therapies, and the real-time monitoring of treatment response, biomarkers have significantly improved patient outcomes and reduced the trial-and-error nature of therapeutic interventions. While challenges in data reproducibility, cost, and regulatory harmonization remain, the convergence of high-throughput omics technologies and artificial intelligence provides a powerful framework for the future of personalized healthcare. Ensuring that these breakthroughs are translated into accessible, equitable clinical practice remains the final, and perhaps most significant, hurdle for the global medical community.

6.14 References

1. The FDA NIH Biomarkers, EndpointS, and other Tools (BEST) resource in neuro-oncology, https://pmc.ncbi.nlm.nih.gov/articles/PMC6071649/
2. The BEST Resource: harmonizing biomarker terminology - FDA, https://www.fda.gov/media/99221/download
3. Personalized medicine: Part 1: Evolution and development into theranostics - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC2957753/
4. The impact of artificial intelligence on biomarker discovery - Portland Press, https://portlandpress.com/emergtoplifesci/article/8/2/89/236530/The-impact-of-artificial-intelligence-on-biomarker
5. About biomarkers and qualification - FDA, https://www.fda.gov/drugs/biomarker-qualification-program/about-biomarkers-and-qualification
6. A review of genomic, transcriptomic, and proteomic applications in edible fungi biology: current status and future directions - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12194055/
7. Artificial intelligence drives advances in multi-omics analysis and precision medicine for sepsis - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12937605/
8. Multi-omics approach in research and drug discovery - Danaher life sciences, https://lifesciences.danaher.com/us/en/library/multi-omics-approach-research-drug-discovery.html
9. Clinical application of peripheral blood biomarkers for solid tumors. https://pmc.ncbi.nlm.nih.gov/articles/PMC12960069/
10. Precision cardiovascular medicine: shifting the innovation paradigm.- Frontiers, https://www.frontiersin.org/journals/science/articles/10.3389/fsci.2025.1474469/full
11. NeuroVoices: Advances and challenges in Alzheimer disease biomarkers. Clinical trials on Alzheimer's disease conference (CTAD). https://www.neurologylive.com/view/neurovoices-nicholas-ashton-advances-challenges-alzheimer-disease-biomarkers
12. Biomarker regulatory considerations - Nordic bioscience, https://www.nordicbioscience.com/why-biomarkers/biomarker-regulatory-considerations/
13. A framework to ethically approach incidental findings in genetic research - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC7745305/

BIOPRESS

14. Precision medicine in low- and middle-income countries - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC9275191/

15. Milestones in personalized medicine: From the ancient time to nowadays-the provocation of COVID-19. https://pmc.ncbi.nlm.nih.gov/articles/PMC7794000/

16. FDA-NIH biomarker working group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-. Glossary. 2016 Jan 28 [Updated 2025 Jan 16]. https://www.ncbi.nlm.nih.gov/books/NBK338448/

17. Making data work: how AI and multi-omics integration are reshaping drug discovery. https://ai.northeastern.edu/news/making-data-work

18. BIOMARKERS related to drug or biotechnology product development - ICH, https://database.ich.org/sites/default/files/ICH%20E16%20Step%204%20presentationWebsite_Oct%202010.pdf

19. FDA Biomarker Guidelines (BEST) - Nordic Bioscience, https://www.nordicbioscience.com/why-biomarkers/fda-biomarker-guidelines-best/

20. Liquid biopsy-derived tumor biomarkers for clinical applications in glioblastoma - MDPI, https://www.mdpi.com/2218-273X/15/5/658

21. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences – PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC6018996/

22. Integrating molecular perspectives: strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics - MDPI, https://www.mdpi.com/2079-7737/13/11/848

23. Genomics, transcriptomics, proteomics, metabolomics and fluxomics for systems-level optimisation. https://www.researchgate.net/publication/397194261_Genomics_Transcriptomics_Proteomics_Metabolomics_and_Fluxomics_for_Systems_Level_Optimization

24. Functional metabolomics: from biomarker discovery to metabolome reprogramming - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC4537470/

25. Applications and challenges of biomarker-based predictive models in proactive health management - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12399543/

26. Why now? - Toward precision medicine - NCBI Bookshelf - NIH, https://www.ncbi.nlm.nih.gov/books/NBK92143/

27. Evolution of precision oncology, personalized medicine, and molecular tumor boards -PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC10894322/

28. E16: Genomic biomarkers related to drug response: context, structure and format of qualification submissions - ICH, https://database.ich.org/sites/default/files/E16_Concept_Paper.pdf

29. ICH E16 Biomarkers related to drug or biotechnology product development: context, structure, and format of qualification submis - FDA, https://www.fda.gov/media/81311/download

30. Types of biomarkers and their applications - Abcam, https://www.abcam.com/en-us/knowledge-center/immunology-and-infectious-disease/types-of-biomarkers-and-their-applications

31. Practical guidance for implementing predictive biomarkers into early phase clinical studies, https://pmc.ncbi.nlm.nih.gov/articles/PMC3819825/

32. Multi-omics strategies for biomarker discovery and application in personalized oncology, https://pmc.ncbi.nlm.nih.gov/articles/PMC12638490/

BIOPRESS

33. Biomarkers in diagnosing and therapeutic monitoring of tuberculosis: a review - PMC, accessed on April 24, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC11299445/

34. Biomarker discovery via in vitro screening - blog - Crown Bioscience, https://blog.crownbio.com/biomarker-discovery-via-in-vitro-screening

35. Biomarker discovery and validation using a combination of in vitro and in vivo studies, https://blog.crownbio.com/biomarker-discovery-and-validation-combining-in-vitro-and-in-vivo-studies

36. Integrative multi-omics pipeline for biomarker discovery in breast cancer using AI-powered bioinformatics tools - Frontiers in Life Sciences Research, https://theeducationjournals.com/index.php/FLSR/article/download/185/193/271

37. Machine learning enhances biomarker discovery: From multi- omics to functional genomics. https://www.researchgate.net/publication/395290977_Machine_learning_enhances_biomarker_discovery_From_multi-_omics_to_functional_genomics

38. Integrative multi-omics characterization and AI-driven biomarker discovery for NSCLC stage III outcomes, https://www.jnjmedicalconnect.com/media/attestation/congresses/oncology/2026/aacr/integrative-multi-omics-characterization-and-ai-driven-biomarker-discovery-for-n.pdf

39. The "Now" and "Next" in precision cardiovascular medicine - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12650113/

40. Biomarkers in infectious diseases - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC6031161/

41. Biomarkers for SARS-CoV-2 infection. A narrative review - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC11978625/

42. Insights into molecular interplay in tuberculosis–COVID-19 Co-Infection via integrated multi-omics strategies - MDPI, https://www.mdpi.com/2571-8800/8/4/41

43. Detection for new biomarkers of tuberculosis infection activity using machine learning methods - MDPI, https://www.mdpi.com/2079-9721/14/2/66

44. Validation framework for in vivo digital measures - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC11751004/

45. Challenges and potential of using digital biomarkers in healthcare and clinical trials - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12996552/

46. Precision-driven biomarker validation: a biotech perspective (Part I) Eureka blog, https://www.criver.com/eureka/precision-driven-biomarker-validation-biotech-perspective-part-i

47. Challenges and opportunities for precision medicine in developing nations. Expert Review of Precision Medicine and Drug Development, 10(1), 1–15. https://doi.org/10.1080/23808993.2025.2505796.

48. Case Studies - Evaluation of biomarkers and surrogate endpoints in chronic disease, https://www.ncbi.nlm.nih.gov/books/NBK220298/

49. Genetic disorder screening: data privacy, consent, and your rights – Accountable. https://www.accountablehq.com/post/genetic-disorder-screening-data-privacy-consent-and-your-rights

50. Privacy in genomics, https://www.genome.gov/about-genomics/policy-issues/Privacy

51. Incidental uses and disclosures - HHS.gov, https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/incidental-uses-and-disclosures/index.html

BIOPRESS

52. ICH E16 Genomic biomarkers related to drug response: context, structure and format of qualification submissions - Scientific guideline | European medicines agency (EMA), https://www.ema.europa.eu/en/ich-e16-genomic-biomarkers-related-drug-response-context-structure-format-qualification-submissions-scientific-guideline

53. Biomarker qualification | European medicines agency (EMA), https://www.ema.europa.eu/en/glossary-terms/biomarker-qualification

54. Biomarkers in medicines development—From discovery to regulatory qualification and beyond - Frontiers, https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.878942/full

55. Unlocking the power of digital biomarkers in clinical trials - ACRP, https://acrpnet.org/2025/08/19/unlocking-the-power-of-digital-biomarkers-in-clinical-trials.

BIOPRESS

Chapter 7

Translational Approaches in Cancer Research: Therapeutic Innovation and Preclinical Evaluation Using Indian Normal and *Cancer Cell Line Models*

Dr.S.Sasikala[1], E. Nivetha[2], R. Geerhanjali[3]

[1]Assistant Professor, Department of Biotechnology, Vinayaka Mission Research Foundation Vinayaka Mission's Kirupanandha Variyar Engineering College, Salem, Tamilnadu, India

[2,3] M.TECH Pharmaceutical Biotechnology, Department of Biotechnological Engineering, Vinayaka Mission Research Foundation, Vinayaka Mission's Kirupanandha Variyar Engineering College, Salem, Tamilnadu, India

---

Abstract

The crucial significance of Indian cell repositories in promoting population-specific medication research is emphasized throughout the chapter. The precision, reproducibility, and translational potential of preclinical and clinical research are enhanced by the biologically relevant models that are provided by native cell lines kept in national repositories and represent India's genetic variety. Additionally, using Indian cell archives encourages independent research and lessens dependency on outside biological resources. In order to gather, maintain, and use drug trial data for successful drug development, this chapter emphasizes the significance of awareness and accountability among scientific professionals, including educators, researchers, physicians, post-doctoral students, nurses, and social workers. It highlights the necessity of ethical research methods and interdisciplinary cooperation to boost biomedical innovation in India.Overall, as a future-focused objective, this chapter promotes greater awareness, infrastructure development, and policy support for Indian cell repositories. Enhancing these repositories will be crucial to the advancement of safe, efficient, and context-specific drug discovery, greatly benefiting India's biomedical research and healthcare system.

Key Words: Genetic Diversity, Translational Research,National Cell Line Repositories and Indian Cancer Cell Lines.

BIOPRESS

7.1 Introduction

The ongoing development of safe and efficient anticancer medications is necessary because cancer continues to be one of the major causes of illness and death worldwide. Because they allow for quick screening, mechanistic research, and toxicity assessment, cell line-based models are essential to clinical drug discovery. Indian Cancer Cell Lines (ICCL) offer special advantages because of their genetic variety, population-specific disease patterns, and significance to domestic medication research, even though a number of international cancer cell lines are commonly used.Transnational oncology research and economical drug discovery in India have been greatly bolstered by the creation of Indian normal and cancer cell banks. [1]

Preclinical research frequently uses a variety of international cancer cell lines, but Indian Cancer Cell Lines (ICCL) have clear advantages because of the country's high genetic heterogeneity, population-specific cancer incidence patterns, and distinct environmental and lifestyle-related risk factors.[2] These features enhance the predicted accuracy of therapy response and resistance mechanisms by enabling ICCLs to more accurately reflect tumor biology pertinent to Indian patients. Furthermore, ICCLs are especially useful for investigating region-specific molecular changes, such as differences in the expression of drug-metabolizing enzymes, tumor suppressor gene mutations, oncogene activation, and epigenetic changes. These variations can have a substantial impact on treatment results, toxicity profiles, and drug sensitivity, highlighting the need of using population-specific cell line models in anticancer drug development pipelines. Transnational oncology research and economical drug discovery in India have been significantly bolstered by the establishment of Indian normal and cancer cell banks. Reproducible research and cooperative investigations across academic, clinical, and industrial sectors are made possible by these repositories, which enable standardized, authenticated, and ethically supplied biological materials. Additionally, having access to native cell lines lessens reliance on expensive imported models and promotes the creation of reasonably priced, regionally appropriate anticancer treatments.[3] Furthermore, precision oncology efforts are improved by integrating ICCLs with cutting-edge technologies including proteomics, transcriptomics, genomes, and high-throughput drug screening platforms.

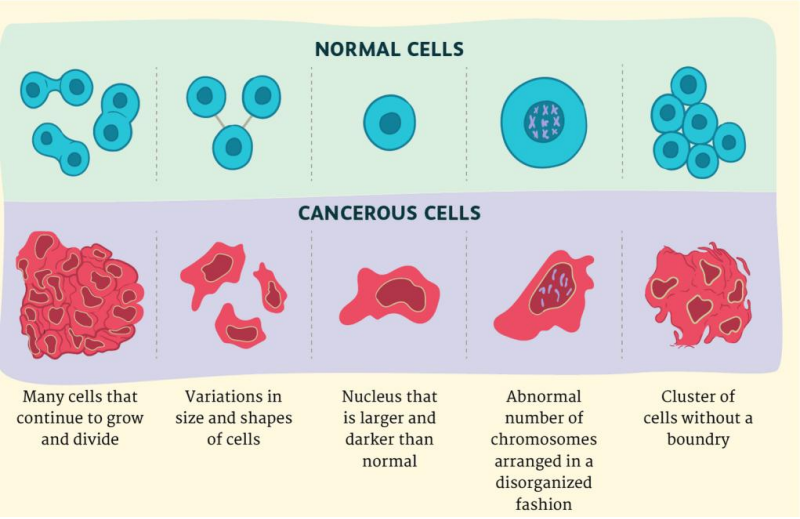

NORMAL CELLS
CANCEROUS CELLS
Many cells that continue to grow and divide
Variations in size and shapes of cells
Nucleus that is larger and darker than normal
Abnormal number of chromosomes arranged in a disorganized fashion
Cluster of cells without a boundry

Figure 7.1: Normal and Cancer Cell Shape and Structure

Indian cancer cell line models play a major role in bridging the gap between preclinical research and clinical translation by facilitating biomarker discovery and customized treatment approaches, thereby enhancing cancer care results both domestically and internationally.[4,5]

7.1.1. Morphological Differences Between Normal and Cancer Cell Lines

The difference between normal and cancer cell lines is mainly based on their growth behavior, genetic stability, and response to regulatory signals. Normal cell lines are derived from healthy tissues and show controlled growth. They require specific growth factors, exhibit contact inhibition (stop dividing when they touch neighboring cells), and have a limited lifespan due to senescence. Their chromosomes are stable, and they maintain normal cell morphology and function. For example, HEK 293 (though transformed, often used as a non-cancer model) and WI-38 are commonly used as normal or near-normal cell models in research. [6] In contrast, cancer cell lines are derived from tumor tissues and show uncontrolled, rapid growth. They lose contact inhibition, can grow without attachment (anchorage-independent growth), and often have mutations in oncogenes or tumor suppressor genes. Cancer cell lines are typically immortal and show abnormal chromosome numbers (haploid). They may also invade surrounding tissues and resist apoptosis (programmed cell death), which are hallmark characteristics of malignant cells. **HeLa cells**, derived from cervical cancer, were first established by George Otto Gay in 1951 and represent one of the earliest and most widely used human cancer cell lines for biomedical research.[7] Similarly, **MCF-7 cells**, derived from human breast adenocarcinoma, were established and characterized by Herbert Soule **and colleagues** in 1973. These cancer cell lines exhibit uncontrolled proliferation, invasive behavior, and resistance to programmed cell death, making them indispensable **in vitro models** for studying tumor biology, cancer progression, and for screening and evaluating anticancer drugs during preclinical drug development.[8]

7.1.2. Significance of Indigenous Normal and Cancer Cell Lines in Drug Discovery

NCCS has been operating a Cancer cell lines are immortalized cells derived from patient tumors that retain key molecular and phenotypic characteristics of cancer. They serve as **in vitro models** for evaluating anticancer agents before animal and clinical studies. High-throughput screening of drug candidates, Understanding molecular mechanisms of drug action Studying drug resistance and sensitivity, Evaluating cytotoxicity and apoptosis, Biomarker identification.[9] Indian cancer cell lines add value by reflecting ethnic-specific genetic variations, which can influence drug response.

7.1.3 Strategic Development of Cell Banking and Biological Data Centers in India

National Cell Repository at **National Centre for Cell Science** (NCCS), Pune

s the **National Cell Repository** since 1986, serving as the primary central repository for mammalian and other animal cell lines in India. It expands, cryopreserves, authenticates, and

distributes cell lines for researchers across academia, government, and industry. The repository provides key services including **cell line authentication (e.g., STR analysis), mycoplasma testing,** and **cell line deposition services** for newly developed indigenous cell lines. Recent annual reports indicate **plans to establish a cGMP-compliant National Mammalian Cell Line Repository & Characterization Facility** under the **National Biopharma Mission** a first-of-its-kind facility meeting international regulatory standards to support industry and academic research. NCCS also conducts **training workshops** on basic and advanced cell culture technologies to build capacity among researchers nationwide. [10]

Species-Specific and Specialized Repositories

ICAR-National Bureau of Fish Genetic Resources **– National Repository for Fish Cell Lines (NRFC):** Established to receive, authenticate, cryopreserve, and distribute **fish-derived cell lines** developed by Indian researchers. It also provides training and access for aquatic cell biology applications (e.g., toxicology, virology, and cellular agriculture research). [11]

Stem Cell and Biobank Initiatives

Institute for Stem Cell Science and Regenerative Medicine **(inStem) Biobank:** A specialized biobank that houses **Indian-origin stem cell lines**, including normal human iPSC lines and patient-derived iPSCs. It supports research into genetic influences on stem cell biology and disease modeling, with plans to expand to include lines for hematological disorders and other conditions.

Broader Data Infrastructure for Life Sciences

Indian Biological Data Center **(IBDC):** Although focused primarily on **biological data,** IBDC serves as a national repository framework for life science data generated in India — an important complement to physical repositories like cell banks by enabling **curated archiving, discovery, and reuse of biological research data.**

Integration with National Biotechnology Policies

National initiatives like the **BioE3 Policy (Biotechnology for Economy, Environment, and Employment**) and the **National Biopharma Mission** have provided strategic direction and funding frameworks that support new repository infrastructure and enable **quality control, compliance, and accessibility** of cell lines for research and biopharma development. [12,13] India has established dedicated centers for the generation, characterization, and distribution of human cancer cell lines, strengthening translational cancer research and drug discovery. The National Centre for Cell Science maintains authenticated Indian cancer cell lines and provides quality-controlled cell culture resources to researchers nationwide. The Advanced Centre for Treatment, Research and Education in Cancer develops clinically relevant patient-derived cancer models that support precision oncology studies. Similarly, the Rajiv Gandhi Centre for Biotechnology contributes to cancer biology research and drug screening platforms. Together, these repositories ensure genetic authentication, contamination-free cultures, and high

reproducibility, thereby enhancing the reliability and impact of biomedical research in India.[14,15]

7.2 Preclinical Drug Discovery Using Indian Oral, Breast, Cervical, and Lung Cancer Cell Lines

Oral Cancer Cell Lines: Oral squamous cell carcinoma is highly prevalent in India due to tobacco use. Indian oral cancer cell lines are crucial for testing anti-oral cancer agents, phytochemicals, and targeted therapies.

Breast Cancer Cell Lines: Indian breast cancer cell lines enable evaluation of hormone-dependent and triple-negative breast cancer therapies, especially for studying drug resistance mechanisms.

Cervical Cancer Cell Lines: Cervical cancer cell lines support screening of HPV-targeted drugs, chemotherapeutic combinations, and radio sensitizers.

Lung and Gastrointestinal Cancer Cell Lines: These lines help evaluate novel chemotherapeutics, kinase inhibitors, and Nano medicine-based delivery systems.

Prostate Cancer Cell Lines: Prostate cancer cell lines are used to evaluate androgen-dependent and androgen-independent therapeutic strategies. These models support studies on hormone therapy resistance, novel anti-androgen drugs, and molecular-targeted treatments.

Ovarian Cancer Cell Lines: Ovarian cancer cell lines are important for testing platinum-based chemotherapeutics, PARP inhibitors, and combination therapies. They also help in understanding chemo résistance and tumor recurrence mechanisms.

Colorectal Cancer Cell Lines: Colorectal cancer models assist in screening targeted therapies such as EGFR and VEGF inhibitors. They are valuable for studying mutation-driven drug responses and personalized medicine approaches.

Hepatocellular Carcinoma (Liver Cancer) Cell Lines: Liver cancer cell lines are used to evaluate anti-proliferative agents, immunotherapeutic, and nanoparticle-based drug delivery systems, especially considering the high burden of viral hepatitis-related liver cancer in India.

Leukemia and Lymphoma Cell Lines: Hematological cancer cell lines support research on targeted therapies, monoclonal antibodies, and kinase inhibitors. These models are essential for understanding molecular pathways involved in blood cancers.

Glioblastoma and Brain Tumor Cell Lines: Brain tumor cell lines are applied in evaluating drugs capable of crossing the blood–brain barrier and studying resistance to radiotherapy and chemotherapy. Together, these diverse Indian cancer cell line models strengthen preclinical drug screening, precision oncology research, and the development of cost-effective therapies tailored to the Indian population.[16,17,18,19]

BIOPRESS

Table: 7.1. Validated Indian Cancer Cell Line Resources in Pharmaceutical Development

Type of Malignancy	Indian-Origin Cell Line Models	Pharmaceutical Screening & Development Utility
Oral Squamous Cell Carcinoma	Patient-derived oral cancer cell lines developed in India	Screening of chemotherapeutic agents, evaluation of phytochemicals, drug resistance studies, apoptosis and EMT analysis (Michael R. Boyd & Kenneth D. Paull *et al*. 1995)
Head and Neck Cancer	Indian head-and-neck squamous carcinoma cell lines	Targeted therapy screening, radio sensitizer evaluation, molecular pathway studies (Ramaiah Nagarajan *et al.*2000)
Breast Cancer	Indian breast cancer cell lines (hormone-positive and aggressive subtypes)	Hormone-based therapy testing, cytotoxicity assays, precision oncology research (Khandekar Jayant D. *et al*.2004 & **Indian Journal of Cancer**, 2013)
Cervical Cancer	Cervical cancer cell lines widely used in Indian laboratories	HPV-targeted drug screening, chemotherapy combinations, radio sensitizer studies (Monk Bradley J. *et al*. 2009)
Lung Cancer	Indian lung carcinoma cell line models	Screening of kinase inhibitors, Nano medicine delivery systems, toxicity profiling (Barretina, J., Caponigro, G., Stransky, N., *et al*. (2012).
Colorectal Cancer	Indian colon cancer cell lines	Evaluation of anticancer compounds, drug metabolism studies, cell cycle analysis (Barretina *et al*., 2012; Shoemaker, 2006; Kumar & Kumar, 2018)
Hepatic Cancer	Indian liver cancer cell lines	Hepatotoxicity assessment, anticancer drug screening, metabolism-linked studies(Gillet, J.-P., Varma, S., & Gottesman, M. M.*et.al* (2013).
Leukemia / Hematological Cancers	Indian leukemia cell lines	Antiproliferative drug testing, signaling pathway analysis, combination therapy studies (Shoemaker, R. H.*et. al.* (2006).
Multiple Cancer Types	Cell line panels maintained by **Advanced Centre for Treatment, Research and Education in Cancer**	High-throughput anticancer drug screening, IC_{50} determination, translational research. (Barretina *et al*., 2012; Shoemaker, 2006; Kumar & Kumar, 2018)

Authenticated Indian Cell Lines	National Repository for Animal Cell Lines, **National Centre for Cell Science**	Standardized cell line supply, reproducibility, quality control in drug discovery (Monk Bradley J. *et al*. 2009)

7.3 Enhancing Drug Efficacy and Safety through Population-Specific In-Vitro Models

Population-specific cell lines are essential tools in modern biomedical research because they accurately represent the **genetic, epigenetic, and environmental characteristics** of a particular population. Human populations differ significantly in terms of genetic polymorphisms, disease susceptibility, metabolism, and response to drugs. Cell lines derived from a specific population therefore provide **more biologically relevant and reliable in-vitro models** than generalized or foreign cell lines.

Genetic variations among populations influence the expression of disease-related genes and signaling pathways. As a result, cell lines established from one population may not fully capture the molecular mechanisms of diseases prevalent in another population. Population-specific cell lines enable researchers to study **region-specific disease patterns**, such as cancers, metabolic disorders, and infectious diseases that show higher incidence or unique progression in certain populations. In pharmacological and toxicological studies, population-specific cell lines are particularly important because drug metabolism and toxicity can vary widely due to genetic differences in enzymes, transporters, and receptors. Using cell lines derived from the target population improves the **predictive accuracy of drug efficacy and safety**, supporting the development of personalized and precision medicine. Additionally, population-specific cell lines reduce scientific bias caused by over-reliance on cell lines derived from limited ethnic backgrounds. They strengthen reproducibility, improve translational relevance, and promote equitable global research. For countries like India, establishing indigenous cell lines supports **cost-effective research, ethical compliance, and national self-reliance,** while ensuring that biomedical innovations are tailored to the healthcare needs of the local population.

7.4 Translational Research Insights from Indian Cancer Cell Line Studies

In recent years, Indian scientists and research institutions have made notable advances in cancer cell line–based studies, contributing to both novel therapeutic strategies and drug discovery knowledge applicable to the Indian population. These studies have enabled the identification of population-specific molecular targets and biomarkers associated with prevalent cancers. Researchers have also explored phytochemical-based anticancer compounds derived from traditional medicinal plants, validating their efficacy using indigenous cell line models. High-throughput drug screening and combination therapy testing have improved understanding of drug resistance mechanisms in Indian tumor subtypes.[20] Furthermore, integration of genomic profiling with in-vitro cell line studies has strengthened precision oncology approaches. Collectively, these efforts are enhancing translational research outcomes and supporting the development of safer, more effective, and regionally relevant anticancer therapies

.**Case Study 1 — Patient-Derived Oral Cancer Cell Lines and Drug Response**

Researchers at the Advanced Centre for Treatment, Research and Education in Cancer have successfully **established and characterized multiple patient-derived squamous cell carcinoma lines** from Indian patients. These cell lines, originating from oral and cutaneous tumors, provide highly relevant in-vitro models for studying tumor biology, epithelial-to-mesenchymal transition (EMT) states, and drug responses in cancers that are especially prevalent in South Asian populations. Such patient-derived models offer **improved translational relevance** when compared with traditional long-established cell lines, enabling more accurate drug sensitivity profiling and discovery of molecular vulnerabilities unique to Indian cancer phenotypes.

Case Study 2 — In-Silico and Cell Line Screening of Indian Phytochemicals against Triple-Negative Breast Cancer

A 2025 study published in *BMC Molecular and Cell Biology* used computational screening of **~300 phytochemicals** against critical oncogenic targets associated with triple-negative breast cancer (TNBC) and then validated molecular interactions using **in-silico docking and simulation methodologies** against relevant cellular targets. Among the phytochemical candidates, **Bayogenin** showed stronger binding affinity to key proteins compared with some established anticancer drugs, suggesting the potential for Indian natural compounds to guide drug design. Although primarily in-silico, such approaches are often followed by **in vitro cell line assays** (e.g., MDA-MB-231 for TNBC) to confirm cytotoxic mechanisms, demonstrating how Indian cell line models integrate with computational drug discovery practices.[21]

Case Study 3 — Anti-Cancer Drug Screening Platforms in India

The Anti-Cancer Drug Screening Facility run by ACTREC provides **in vitro screening across over 50 human tumor cell lines,** enabling academic and industrial partners to evaluate large chemical libraries for cytotoxic activity. This facility has successfully identified compounds with potent activity against oral, hepatic, and colorectal cancer cell lines and has facilitated in vivo follow-ups using patient-derived xenografts in collaboration with partner institutions. This infrastructure represents an **Indian national resource** accelerating early drug discovery, particularly for orally prevalent carcinomas like oral cavity and head-and-neck cancers.[22]

Case Study 4 — Plant-Derived Therapeutic Leads Tested on Cell Lines

In Tamil Nadu, a funded research project examined extracts of *Caesalpinia bonducella* for its effects on **breast and colon cancer cell lines (e.g., MCF-7 and HT-29)**. This translational project — supported by state research grants — aims to bridge ethnobotanical leads with modern cell line cytotoxicity assays, a model for how indigenous knowledge can contribute to novel therapeutic identification via cell line research.[23]

7.5 Importance of Indian Cell Line Establishment

Genetic and Population Relevance: Indian cell lines reflect the vast genetic diversity of the Indian population, enabling accurate disease modeling and pharmacogenomics studies. They help researchers investigate population-specific disease patterns and understand variations in drug response compared to Western populations. **Reducing Dependence on Imported Cell Lines:** The establishment of indigenous cell lines lowers research costs, minimizes regulatory and import-related delays, and ensures uninterrupted availability of biological materials for research and diagnostics within the country. **Disease-Specific Research (Cancer & Infectious Diseases):** Indian cell lines are crucial for studying diseases highly prevalent in India, including oral, cervical, and breast cancers, diabetes, tuberculosis, and other infectious and genetic disorders. This improves the development of relevant disease models and enhances the accuracy of experimental outcomes. **Drug Discovery, Toxicology, and Vaccine Development:** Indigenous cell lines strengthen preclinical drug screening, toxicity assessment, and vaccine research. They improve translational relevance, reduce dependency on animal models, and promote more ethical and cost-effective research practices. **Quality Control, Authentication, and Biosafety:** Establishing well-characterized and authenticated Indian cell lines prevents cross-contamination, ensures reproducibility, and maintains biosafety standards. Proper genetic authentication enhances the reliability and credibility of scientific findings. **Capacity Building and Indigenous Innovation:** National cell line repositories support academic institutions, biotechnology startups, and pharmaceutical industries, aligning with initiatives such as Atmanirbhar Bharat. These facilities also provide hands-on training for students, researchers, and laboratory professionals, strengthening technical expertise in cell culture and molecular biology. **Ethical and Regulatory Compliance:** Indian cell line establishment follows national ethical guidelines and regulatory frameworks, ensuring responsible sourcing and use of human and animal biological materials. **Computational and Cross-Disciplinary Support:** In addition to experimental platforms, Indian researchers contribute to bioinformatics databases and integrative computational tools that complement cell line research. Curated resources, such as tumor-homing peptide databases linked to specific cancer cell lines, assist in targeted drug screening workflows, reduce experimental costs, and improve hit-validation success rates in drug discovery programs.[24,25]

7.6 Indian vs. Global Cancer Cell Lines: A Comparative Perspective

Morphological Differences: Indian vs. Global Cancer Cell Lines

While cancer cell lines generally share core malignant features, morphological variations can arise due to genetic background, tumor origin, and adaptation to local environmental factors. These differences may influence cellular architecture, growth behavior, and differentiation status observed under in-vitro conditions. Variations in stromal interactions, prior exposure to carcinogens (such as tobacco-related compounds), and viral associations can further shape morphological characteristics. Additionally, passage number and culture conditions may contribute to structural changes over time, leading to phenotypic drift. Therefore, careful

morphological characterization, along with molecular validation, is essential to ensure that cell lines accurately represent the biological features of the original tumor.

Cellular Shape and Uniformity: Indian Cancer Cell Lines: May show morphological patterns reflecting region-specific tumor biology, including variations in epithelial or mesenchymal features. In certain oral and cervical cancer models, higher cellular pleomorphism (variation in size and shape) has been reported, consistent with aggressive local tumor characteristics. **Global Cancer Cell Lines:** Often well characterized and standardized after decades of continuous passaging. Many exhibit relatively stable morphology due to long-term adaptation to in-vitro conditions.

Nuclear Characteristics: **Indian Cell Lines:** May demonstrate prominent nucleoli and variable nuclear–cytoplasmic ratios depending on tumor subtype and grade, particularly in cancers prevalent in India such as oral squamous cell carcinoma. **Global Cell Lines:** Classic lines such as HeLa and MCF-7 show consistent enlarged nuclei, hyperchromasia, and stable nuclear morphology due to extensive characterization.

Growth Pattern in Culture: **Indian Cancer Cell Lines:** Some recently developed patient-derived lines may retain tumor heterogeneity and display mixed growth patterns (adherent clusters with partial loss of contact inhibition). **Global Cancer Cell Lines:** Many long-established lines exhibit highly reproducible monolayer growth or suspension growth patterns (e.g., leukemia models), optimized for laboratory scalability.

Global Cell Lines: Long-term culturing can sometimes lead to phenotypic drift, resulting in reduced resemblance to the original tumor tissue.

Tumor Heterogeneity Representation: **Indian Cell Lines:** Often developed to represent population-specific cancers (oral, cervical, tobacco-related malignancies), potentially preserving distinct morphological traits. **Global Cell Lines:** While diverse, many widely used lines originate predominantly from Western populations, which may not fully reflect the heterogeneity seen in Indian patients. Overall, morphological differences are influenced more by tumor origin, genetic background, and passage history rather than nationality alone. However, Indian cancer cell lines may better capture morphological diversity relevant to region-specific disease patterns, thereby enhancing translational relevance in local drug discovery research.

Global cancer cell lines widely used in North America and Europe have been fundamental to anticancer drug discovery, enabling standardized high-throughput screening and mechanistic research. However, they often represent limited ethnic diversity and may not fully capture population-specific disease heterogeneity. In contrast, Indian cancer cell lines developed by institutions such as the Advanced Centre for Treatment, Research and Education in Cancer and curated by the National Centre for Cell Science better reflect the genetic, environmental, and lifestyle-related factors relevant to the Indian population. They are particularly valuable for studying region-specific cancers like oral and head-and-neck malignancies. While global

models offer extensive molecular annotation and reproducibility, Indian cell lines enhance translational relevance, improve drug response prediction, and support cost-effective precision oncology. The combined use of both models strengthens preclinical research and clinical translation.

Table 7.2: Evaluation of Cancer Cell Line Models in Drug Development: Cross-Population Perspective.

Assessment Factors	Established Global Cell Line Models	Population-Specific Indian Cell Lines
Geographic origin	Primarily developed and standardized in North America and Europe	Developed within India through indigenous research programs
Genetic & ethnic representation	Represent limited ethnic and genetic diversity	Reflect genetic, environmental, and lifestyle diversity of the Indian population
Role in drug discovery	Enable standardized high-throughput screening and mechanistic studies	Provide population-specific insights for drug response and efficacy
Cancer types represented	Broad tumor coverage but limited representation of region-specific cancers	Strong representation of region-specific cancers such as oral and head-and-neck malignancies
Molecular annotation	Extensive molecular characterization and long-term reproducibility	Increasing molecular profiling with growing clinical relevance
Translational relevance	Useful for global benchmarking and early discovery	High translational relevance for Indian and South Asian populations
Cost and accessibility	Often expensive and dependent on international repositories	Cost-effective and locally accessible for Indian researchers
Key supporting institutions	International cancer research repositories and consortia	Advanced Centre for Treatment, Research and Education in Cancer and National Centre for Cell Science
Overall contribution	Provides historical consistency and global comparability	Enhances precision oncology and locally relevant drug discovery
Best research strategy	Effective when used as global reference models	Most powerful when used alongside global cell lines

7.7 Application of Indian Cancer Cell Lines in Drug Discovery Pipeline

Indian cancer cell lines play a crucial role across multiple stages of the drug discovery pipeline, from target identification and validation to preclinical drug screening and mechanism-of-action

studies. These cell lines provide population-specific biological models that reflect India's unique genetic, environmental, and lifestyle diversity, thereby improving the relevance and predictive accuracy of anticancer drug research. In early-stage discovery, they enable high-throughput screening of novel compounds and help identify molecular targets specific to Indian patient populations. During preclinical evaluation, Indian cancer cell lines support toxicity assessment, biomarker discovery, and resistance mechanism studies, enhancing translational reliability. Their integration into drug development strategies not only reduces dependence on foreign cell models but also accelerates the development of safer, more effective, and precision-based cancer therapeutics tailored to the Indian population.

7.8 Innovation Through Indigenous Resources: The Vision for India's Cell Repository

Building a sustainable Indian cell line repository requires coordinated efforts from multiple stakeholders. Government and national funding agencies such as the Indian Council of Medical Research, Department of Biotechnology, and Council of Scientific and Industrial Research provide policy direction, ethical regulation, infrastructure funding, and translational research support. National institutes and repositories, including the National Centre for Cell Science, serve as the backbone for cell line derivation, authentication, cryopreservation, and distribution. Universities and academic researchers contribute by developing disease-specific cell lines and training skilled professionals, while hospitals supply ethically consented patient samples and linked clinical data. Biotechnology and pharmaceutical industries strengthen translation, funding, and large-scale drug screening through public–private partnerships. Ethical committees ensure regulatory compliance and donor protection, and policymakers align repository development with national health priorities and initiatives such as Atmanirbhar Bharat, ensuring scientific sustainability and global credibility.

7.9 Awareness in Action: Building India's Cell Line Legacy"

Creating awareness about Indian cell line repository establishment requires a multi-pronged approach targeting students, researchers, policymakers, and industry stakeholders. At the academic level, integrating topics on Indian cell lines into UG and PG curricula, organizing guest lectures, webinars, FDPs, seminars, and journal clubs can cultivate early scientific responsibility and interest. Hands-on training workshops, internships, and programs on cell culture, authentication, and ethics—facilitated by institutions like the National Centre for Cell Science—enhance practical understanding and acceptance. Government and policy-level initiatives by agencies such as the Department of Biotechnology and Indian Council of Medical Research, including national campaigns, funding calls, and policy guidelines, promote self-reliance and structured repository development. Scientific conferences, symposiums, and dedicated sessions provide platforms for collaboration and visibility across the research community. Industry–academia partnerships, highlighting benefits like cost reduction and accelerated R&D, ensure sustainability and wider adoption. Ethical and regulatory education on informed consent, biosafety, and simplified SOPs builds trust and confidence among researchers. Digital outreach through dedicated websites, social media campaigns, explainer videos, and online databases improves accessibility and usage. Finally, recognition and

incentives—awards, funding support, and academic citations—motivate researchers to contribute indigenous cell lines, fostering a culture of collaboration and national scientific advancement.

7.10 Tissue Donation for Indian Cell Line Repositories: Engaging Participants and Recognizing Contributions

Ensuring effective and ethical tissue donation for Indian cell line repository development requires clear, respectful, and culturally sensitive communication with participants. Information should be conveyed in simple, non-technical language, ideally in the participant's local language, so they understand what a cell line is, the importance of their tissue for population-specific research, and that donation will not affect their treatment or health. One-to-one explanations by trained medical staff, supplemented with brochures, charts, and short videos, are most effective. Participation must be emphasized as voluntary, with the freedom to refuse or withdraw consent at any time, fostering trust and confidence. Institutional Ethics Committees should also inform participants about ethical safeguards, including confidentiality, secure storage, controlled use of samples, and approvals in accordance with Indian Council of Medical Research guidelines, reducing fear and misconceptions. Participants may also receive health- or knowledge-based benefits, including free basic check-ups, access to health education programs, and priority counseling, which are considered supportive rather than inducements. Long-term engagement through updates on research outcomes and success stories fosters goodwill and a sense of belonging to the national scientific mission. Communication should respect cultural, religious, and emotional considerations, with counselors or community health workers aiding in addressing concerns and improving acceptance across India's diverse population.

7.11 Advantages of Using Indian Cancer Cell Lines

Indian cancer cell lines offer genetic relevance to the local population, are cost-effective and easily accessible, and support indigenous drug development. They enable research on region-specific cancers, reduce dependence on imported models, and enhance reproducibility of studies. These cell lines facilitate precision medicine, accelerate preclinical and translational research, strengthen national scientific infrastructure, and improve the predictive value of clinical outcomes while promoting ethical and regulated use of human tissues.

7.12 Challenges and Limitations

Despite their advantages, Indian cancer cell lines face several challenges that limit their broader utility. The availability of well-characterized lines, particularly for rare or region-specific cancers, remains limited. Long-term culture can lead to genetic drift, altering cell behavior and reducing reproducibility. Variability in culture, authentication, and storage protocols across laboratories further complicates standardization. Integration with advanced models such as 3D cultures, organoids, and co-culture systems is still limited, and resource constraints—including costs of maintenance and cryopreservation—pose additional hurdles. Ethical and regulatory

complexities, such as obtaining informed consent and complying with national guidelines, require careful attention. Awareness among researchers is often low, while limited funding and policy support impede the expansion and sustainability of repositories. Incomplete genomic and molecular characterization, along with the risk of contamination or misidentification, underscores the need for continuous authentication, robust quality control, and strategic investment in infrastructure to maximize the research potential of Indian cancer cell lines.

7.13 Vision 2030: Indian Cell Line Bank Repository for Precision Oncology and Cancer Cure

The establishment of an indigenous Indian cell line bank repository is envisioned as a cornerstone initiative in India's mission to transform cancer research, precision oncology, and drug discovery by the year 2030. This repository is not merely a biological resource center but a national scientific infrastructure designed to enable population-specific, data-driven, and patient-relevant cancer research tailored to the genetic and environmental diversity of India.

Cancer in India exhibits distinct epidemiological patterns, driven by region-specific risk factors such as tobacco consumption, HPV infection, dietary habits, pollution, and socio-economic determinants. A centralized repository of Indian-origin human cancer cell lines—derived from oral, breast, cervical, lung, gastrointestinal, prostate, ovarian, liver, hematological, and brain tumors—will provide biologically accurate models to decode these unique disease signatures. These models will form the foundation for discovering therapies that are more effective, less toxic, and better suited to Indian patients.

By 2030, the Indian cell line bank aims to serve as a catalyst for precision oncology, enabling the stratification of patients based on molecular profiles and treatment responsiveness. Integration of cell line data with genomics, proteomics, metabolomics, and artificial intelligence will allow rapid identification of drug targets, prediction of therapeutic outcomes, and optimization of personalized treatment regimens. Such an approach has the potential to significantly reduce trial-and-error treatment strategies and improve survival rates across multiple cancer types.

The repository will also accelerate the development and validation of next-generation therapies, including targeted drugs, immunotherapies, nanomedicine-based delivery systems, and combination treatment strategies. Importantly, Indian cell line models will support the evaluation of traditional medicine-derived compounds and phytochemicals, scientifically bridging India's rich heritage of natural products with modern drug discovery pipelines. From a national perspective, the Indian cell line bank will contribute to scientific self-reliance, reduce dependency on imported biological resources, and lower the cost of drug development. It will foster large-scale collaboration between academic institutions, hospitals, biotechnology startups, and pharmaceutical industries, creating an innovation ecosystem aligned with the vision of affordable and accessible cancer care. Ultimately, the Vision 2030 Indian Cell Line Bank Repository aspires to move cancer research from generalized treatment to truly personalized therapy, where the right drug reaches the right patient at the right time. By

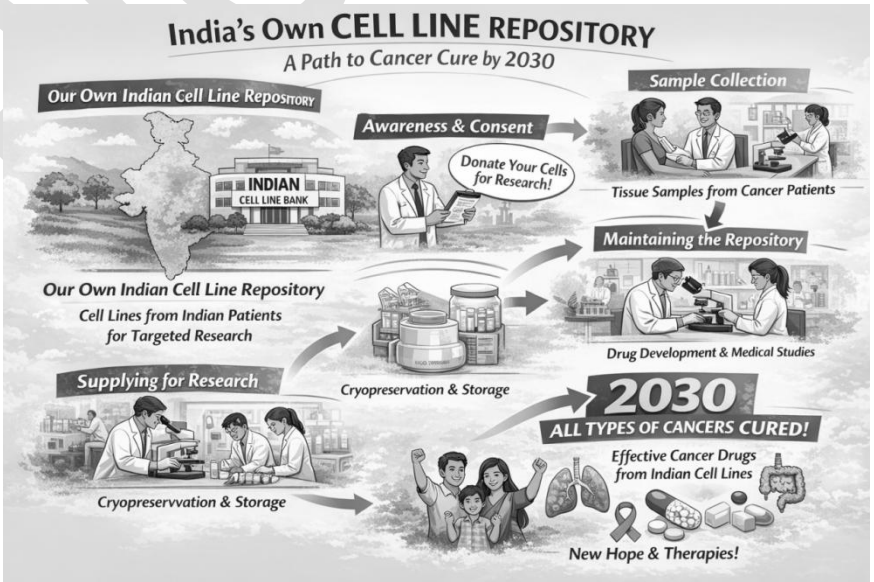
anchoring drug discovery in Indian-specific cellular models, this initiative holds the promise of transforming cancer from a life-threatening disease into a manageable and, in many cases, curable condition for future generations.

7.14 Future Perspectives

The future of Indian cancer cell lines in drug discovery lies in integrating omics technologies, developing 3D spheroids and organoids, and leveraging AI-driven drug screening platforms. Linking cell line data to clinical outcomes will enhance translational relevance and precision medicine. Expanding well-annotated biobanks, incorporating CRISPR and gene-editing tools, standardizing culture protocols, and focusing on rare cancers will strengthen research quality. Collaboration with industry and international consortia, along with policy support, sustainable funding, and open-access data sharing, will ensure India's cell line resources contribute effectively to global oncology drug development.

7.15 Conclusion

Indian cancer cell lines serve as vital tools in effective drug discovery by providing population-specific relevance and cost-efficient research platforms. Their integration into the drug development pipeline strengthens translational research, accelerates anticancer drug innovation, and supports precision medicine initiatives in India. Promoting awareness among educators, researchers, clinicians, post-doctoral fellows, nurses, and social workers is essential to advance responsible drug trial practices and ethical data management. Indian cell repositories, such as the National Centre for Cell Science, play a critical role in preserving indigenous cell lines that reflect the nation's genetic and environmental diversity. Strengthening these repositories enhances research reproducibility, reduces reliance on foreign models, and supports the development of safer, population-specific, and more effective therapeutics, reinforcing India's self-reliance in biomedical research.


India's Own CELL LINE REPOSITORY
A Path to Cancer Cure by 2030
Our Own Indian Cell Line Repository
INDIAN
CELL LINE BANK
Awareness & Consent
Donate Your Cells for Research!
Sample Collection
Tissue Samples from Cancer Patients
Maintaining the Repository
Drug Development & Medical Studies
Our Own Indian Cell Line Repository
Cell Lines from Indian Patients for Targeted Research
Cryopreservation & Storage
Supplying for Research
Cryopreservvation & Storage
2030
ALL TYPES OF CANCERS CURED!
Effective Cancer Drugs from Indian Cell Lines
New Hope & Therapies!

Advancing drug discovery demands collective responsibility from educators, researchers, clinicians, and allied health professionals, grounded in ethical trial practices, transparent data management, and regulatory compliance. Interdisciplinary collaboration enhances scientific rigor, reproducibility, and effective translation of research into clinical practice. Indian cell repositories, including the National Centre for Cell Science and the National Centre for Cell Science - National Repository for Cell Lines, are central to building a self-reliant research ecosystem. Strengthening indigenous, population-specific cell lines reduces dependence on foreign models and improves the reliability and predictive value of preclinical and clinical studies. With sustained policy support and infrastructure development, India can advance toward ethical, sustainable, and population-centered. **"From Indian Cells to Indian Cures – A Nation's Step Toward Cancer-Free Tomorrow".**

7.16 References

1. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., ... & Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature, 483(7391), 603-607.
2. Gillet, J. P., Varma, S., & Gottesman, M. M. (2013). The clinical relevance of cancer cell lines. Journal of the National Cancer Institute, 105(7), 452-458.
3. Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. Nature Reviews Cancer, 6(10), 813-823.
4. Advani, D., Sharma, S., Kumari, S., Ambasta, R. K., & Kumar, P. (2022). Precision oncology, signaling, and anticancer agents in cancer therapeutics. Anti-Cancer Agents in Medicinal Chemistry-Anti-Cancer Agents), 22(3), 433-468.
5. Roy, R., De Sarkar, N., Ghose, S., Paul, R. R., Pal, M., Bhattacharya, C., ... & Roy, B. (2014). Genetic variations at microRNA and processing genes and risk of oral cancer. Tumor Biology, 35(4), 3409-3414.
6. George, A., Nair, S., Prabhash, K., Thakur, S., Gera, P., Singh, A., ... & Kode, J. A. (2026). Generation and characterization of early stage oral cancer cell line of buccal mucosa of Indian origin. *Human Cell*, *39*(1), 21.
7. Warawdekar, U. M., & Kowtal, P. (2015). Molecular pathways to therapeutics: Paradigms and challenges in oncology meeting report: Carcinogenesis 2015. *Journal of Carcinogenesis*, *14*, 4.
8. Gacche, R., Birge, R. B., & Chandra, D. (2024). Biology beyond boundaries: Mitochondrial insights, computational breakthroughs, and clinical transformations: International conference hosted by Savitribai Phule Pune University, Pune, India, January 29–31, 2024. *Cell Communication and Signaling*, *22*(1), 406.
9. Sankaranarayanan, P., Abhinand, P. A., Manikandan, M., & Ghosh, A. (2025). Molecular docking and MD simulation approach to identify potential phytochemical lead molecule against triple negative breast cancer. *F1000Research*, *13*, 1271.
10. Joshi, P., Bane, S., Chaturvedi, P., Gera, P., & Waghmare, S. K. (2025). Establishment and characterization of patient-derived tongue squamous cell carcinoma cell lines: Establishment and characterization of tongue carcinoma cell lines. *Human Cell*, *38*(4), 102.
11. Matters, A. (2015). Biochemists of India (ACBICON 2015). *Ind J Clin Biochem*, *30*(1), S1-S131.
12. Singh, T. A. (2015). Abstracts of 42nd National Conference of Association of Clinical Biochemists of India (ACBICON 2015). *Indian Journal of Clinical Biochemistry*.
13. Sharma, S., Kumar, C., Kushwaha, H., Jha, S. K., Chawla, S., Sharma, A., ... & Huddar, V. (2025). Advancing anticancer drug development: Overcoming challenges and exploring new therapeutic strategies. *Ayush Journal of Integrative Oncology*, *2*(1), 8-27.
14. Kumarasamy, R. V., Natarajan, P. M., Mathivanan, I., Balasubramaniam, M., SN, S., Prabhu, D., ... & Mironescu, I. D. (2025). Sustainable biogenic synthesis of silver nanoparticles for oral cancer: a comprehensive review. *Frontiers in Nanotechnology*, *7*, 1648900.

15. James, C., Whitehead, A., Plummer, J. T., Thompson, R., & Badal, S. (2024). Failure to progress: breast and prostate cancer cell lines in developing targeted therapies. *Cancer and Metastasis Reviews*, *43*(4), 1529-1548.
16. Masters, J. R. (2000). Human cancer cell lines: fact and fantasy. *Nature reviews Molecular cell biology*, *1*(3), 233-236.
17. Freshney, R. I. (2015). *Culture of animal cells: a manual of basic technique and specialized applications*. John Wiley & Sons.
18. Masters, J. (Ed.). (2012). *Human cancer in primary culture, a handbook*. Springer Science & Business Media.
19. Kerr, J. F., Wyllie, A. H., & Currie, A. R. (1972). Apoptosis: a basic biological phenomenon with wideranging implications in tissue kinetics. *British journal of cancer*, *26*(4), 239-257.
20. Alberts, B. (2002). Molecular biology of the cell 4th edition.
21. Godoy, P., Hewitt, N. J., Albrecht, U., Andersen, M. E., Ansari, N., Bhattacharya, S., & Hengstler, J. G. (2013). Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. *Archives of toxicology*, *87*(8), 1315-1530.
22. Vicente, E., Lesniewski, M., Newman, D., Vujaskovic, Z., & Jackson, I. L. (2022). Best practices for authentication of cell lines to ensure data reproducibility and integrity. *Radiation Research*, *197*(3), 209-217.
23. Nandikolmath, M. V. (2023). *Principles of Animal Cell Technology: A Practical Approach (Volume: 1)*. Shineeks Publishers.

BIOPRESS

Chapter 8

Regulatory Considerations in New Drug Approval Pathways

Ashwini Badhe

Swalife Foundation, Pune, India

Abstract

The pharmaceutical industry operates within one of the most stringent and intricate regulatory environments in the global economy, primarily because the products it generates directly impact human life and public safety. This report provides a comprehensive examination of the regulatory considerations that govern the lifecycle of new drug approval, from the nascent stages of drug discovery through to post-marketing surveillance. Central to this analysis is the critical role of regulatory frameworks in ensuring that every medicinal product introduced to the market adheres to the highest standards of safety, efficacy, and quality. The discussion explores the divergent yet increasingly harmonized approval pathways across major jurisdictions, including the United States, Europe, Japan, and India. By analyzing the functions of key agencies such as the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), and the Pharmaceuticals and Medical Devices Agency (PMDA), the report highlights the mechanisms that facilitate global market access while maintaining rigorous oversight.

Special attention is given to the International Council for Harmonisation (ICH) and its success in standardizing technical requirements through the Common Technical Document (CTD). The report further investigates the burgeoning role of transformative technologies, such as Artificial Intelligence (AI) and Machine Learning (ML), in lead optimization and biomarker identification, which are reshaping the evidentiary standards for clinical trials. Furthermore, it addresses the evolving use of Real-World Evidence (RWE) and multi-omics integration in modernizing regulatory decision-making. Despite these innovations, the path to drug approval remains fraught with challenges, including significant costs, regulatory delays, and the ethical complexities of data privacy and global harmonization. The report concludes with a forward-

BIOPRESS

looking perspective on the digital transformation of regulatory science and the emergence of personalized medicine regulations, advocating for continued innovation and international collaboration to ensure timely patient access to safe and effective therapies.

Keywords

Drug approval, Regulatory pathways, Clinical trials, Investigational New Drug (IND), New Drug Application (NDA), Pharmacovigilance, ICH guidelines, Common Technical Document (CTD), Chemistry, Manufacturing, and Controls (CMC), Real-World Evidence (RWE), Artificial Intelligence (AI).

8.1 Introduction

The regulation of pharmaceutical products is not merely a bureaucratic hurdle but a fundamental scientific and ethical imperative. Historically, drug regulation has evolved from reactive measures following public health crises to a proactive, science-based discipline that oversees the entire development lifecycle. The primary objective of these regulatory systems is the assurance of safety, efficacy, and quality.[1] Safety involves the identification and mitigation of toxicities, while efficacy refers to the demonstrable therapeutic benefit of a drug for its intended indication. Quality, perhaps the most technical of the three, ensures that every batch of a drug is manufactured consistently to predetermined specifications.

The global drug development process is an arduous journey, typically spanning over a decade and costing billions of dollars. It begins with basic research and target identification, moves through preclinical testing to evaluate biological activity and toxicity, and culminates in human clinical trials.[14] Each of these stages is governed by a complex web of regulatory requirements that differ by jurisdiction, making the need for regulatory compliance a strategic priority for any pharmaceutical sponsor. Navigating these requirements effectively is essential for minimizing the "drug lag"—the time between a drug's first global approval and its availability in other regions.

The scope of this report encompasses the entire spectrum of regulatory considerations. It details the standardized processes established by the ICH, the specific procedural tracks utilized by major global health authorities, and the technical requirements for manufacturing and submission. Furthermore, it addresses the modern shift toward expedited approval pathways and the integration of digital tools and real-world data, which are increasingly critical in the age of precision medicine.[16] By synthesizing insights from current regulatory reports and

scientific literature, the analysis provides a nuanced understanding of how regulatory science is adapting to the challenges of the 21st century.

8.2 Overview of Drug Development Process

The pathway from a molecular hit to a marketed drug is a multi-phase process characterized by high attrition and rigorous oversight. Understanding the progression of a candidate molecule through this pipeline is essential for appreciating the regulatory milestones that must be met.

8.3 Discovery and Preclinical Studies

The discovery phase involves identifying a biological target (such as a protein or enzyme) associated with a disease and finding "hits"—molecules that interact with that target. This often involves high-throughput screening (HTS), which can evaluate thousands of compounds in a matter of days. Once a hit is confirmed, it enters "lead optimization," where medicinal chemists refine its chemical structure to improve potency and selectivity while addressing its absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile.

Preclinical studies are conducted in vitro (cell-based) and in vivo (animal-based) to characterize the drug's safety and biological activity before human exposure. These studies must be conducted under Good Laboratory Practice (GLP) to ensure the integrity and reliability of the data. The goal is to establish a starting dose for human trials and to identify potential target organ toxicities.

8.4 Clinical Trials (Phase I–IV)

Clinical development is the most visible and expensive phase of drug development, typically divided into four distinct phases:

- **Phase I:** These trials are the first in-human studies, usually involving a small group of healthy volunteers (20–100). The primary focus is safety, tolerability, and determining the pharmacokinetic (PK) and pharmacodynamic (PD) profile.
- **Phase II:** The drug is tested in a larger group of patients (100–300) with the target disease. The objective is to evaluate efficacy (proof of concept) and to further monitor safety and dose-response relationships.
- **Phase III:** These are large-scale, multi-center trials involving thousands of patients. They are designed to provide definitive evidence of efficacy and safety compared to the

current standard of care or a placebo. The data from Phase III trials form the core of the marketing authorization application.

- **Phase IV:** Also known as post-marketing surveillance, these trials occur after the drug has been approved. They monitor long-term safety, rare adverse events, and efficacy in diverse real-world populations.

8.5 Regulatory Submission and Approval

Upon successful completion of Phase III, the sponsor compiles all accumulated data—preclinical, clinical, and manufacturing—into a comprehensive dossier. In the U.S., this is the New Drug Application (NDA) or Biologics License Application (BLA); in Europe, it is the Marketing Authorization Application (MAA). Regulatory authorities review the application to determine if the drug's benefits outweigh its risks for the proposed indication.

8.6 Flow Diagram of the Drug Development and Approval Pathway

The following flow diagram illustrates the sequential progression and the regulatory intersections of the drug development process:

Table 8.1: Differenet phases of clinical trials

Stage	Activity	Key Regulatory Milestone
Discovery	Target ID, Hit Discovery, Lead Opt	Internal Go/No-Go Decision
Preclinical	GLP Safety, Pharmacology, ADMET	IND / CTN Submission
Phase I	Human Safety, PK/PD, Dose Escalation	Institutional Review Board (IRB) Approval
Phase II	Efficacy Proof-of-Concept, Dose Selection	End-of-Phase II Meeting with Agency

Phase III	Pivotal Efficacy & Safety, Scale-up	Preparation of CTD / eCTD Dossier
Review	Agency Scrutiny of NDA / MAA	Marketing Authorization Approval
Post-Market	Phase IV, Pharmacovigilance, RWE	Risk Management Plan (RMP) Execution

8.7 Key Regulatory Authorities

Drug regulation is governed by national and regional agencies that enforce laws and guidelines to protect public health. While their ultimate goals are identical, their structures and procedures vary.

8.7.1 U.S. Food and Drug Administration (USA)

The FDA is often considered the preeminent regulatory authority globally. It operates under the Department of Health and Human Services and is responsible for regulating human and veterinary drugs, biologics, and medical devices. Within the FDA, the **Center for Drug Evaluation and Research (CDER)** oversees the approval of small-molecule drugs, while the **Center for Biologics Evaluation and Research (CBER)** handles biologics, including vaccines and gene therapies. The FDA is known for its science-driven approach and transparent review processes, often setting benchmarks that influence global standards.

8.7.2 European Medicines Agency (Europe)

The EMA is a decentralized agency of the European Union (EU) that coordinates the evaluation and monitoring of medicines across the member states. It utilizes a network of thousands of experts from national competent authorities. The EMA's **Committee for Medicinal Products for Human Use (CHMP)** provides scientific opinions on drug applications. Most innovative medicines are approved through the **Centralized Procedure**, which results in a single authorization valid in all EU and EEA countries.

8.7.3 Central Drugs Standard Control Organization (India)

BIOPRESS

The CDSCO is India's national regulatory body, headed by the **Drug Controller General of India (DCGI)**. Under the **New Drugs and Clinical Trials (NDCT) Rules, 2019**, the CDSCO has undergone substantial reforms to align its processes with international standards. It utilizes **Subject Expert Committees (SECs)** to evaluate technical data and provide recommendations on clinical trial approvals and new drug marketing authorizations.

8.7.4 Pharmaceuticals and Medical Devices Agency (Japan)

The PMDA is the Japanese agency responsible for the scientific review of drugs and medical devices. It works in close collaboration with the **Ministry of Health, Labour and Welfare (MHLW)**, which grants the final marketing authorization. Japan has dramatically reduced its review times over the last two decades, moving from a position of "drug lag" to being a leader in expedited pathways like **Sakigake**.

8.7.5 World Health Organization (Global Guidance)

While the WHO does not grant marketing authorizations, it provides essential global guidance and sets international standards for drug quality and safety. Its **Prequalification Programme** is vital for ensuring that medicines supplied by international procurement agencies for use in developing countries meet high quality standards.

Table 8.2: Different regulatory bodies of various countries

Agency	Jurisdiction	Key Role / Strength
FDA	United States	Science-driven, transparent, sets global benchmarks.
EMA	European Union	Centralized evaluation, focus on long-term safety.
CDSCO	India	Rapidly evolving reforms, focus on affordable generics & innovative trials.
PMDA	Japan	Efficiency, leader in expedited innovative pathways.

WHO	Global	Standards for quality, prequalification for developing nations.

8.8 Regulatory Frameworks and Guidelines

The harmonization of regulatory requirements is primarily facilitated by the **International Council for Harmonisation (ICH)**, which develops technical guidelines that are adopted by regulatory authorities worldwide.

8.8.1 ICH Guidelines

The ICH guidelines are organized into four main categories: Quality (Q), Safety (S), Efficacy (E), and Multidisciplinary (M).

- **Quality Guidelines (Q):** These cover CMC aspects, including stability testing (Q1), analytical validation (Q2), impurities (Q3), and Good Manufacturing Practice (GMP) for APIs (Q7). The Q8 guideline introduced the concept of **Quality by Design (QbD)**, focusing on building quality into the product through understanding the manufacturing process.
- **Safety Guidelines (S):** These address non-clinical testing to identify potential risks such as carcinogenicity, genotoxicity, and reproductive toxicity.
- **Efficacy Guidelines (E):** These govern the design, conduct, and reporting of clinical trials. Key guidelines include **Good Clinical Practice (GCP, E6)** and those related to pediatric studies (E11) and pharmacovigilance (E2C).
- **Multidisciplinary Guidelines (M):** These include the **Common Technical Document (CTD, M4)** and the **Medical Dictionary for Regulatory Activities (MedDRA, M1)**, which provide a standardized medical terminology for regulatory communication.

8.8.2 Good Practices (GxP)

- **Good Laboratory Practice (GLP):** A quality system concerned with the organizational process and conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived, and reported.

BIOPRESS

- **Good Clinical Practice (GCP):** An international ethical and scientific quality standard for clinical trials. It ensures that the rights, safety, and well-being of trial subjects are protected and that the clinical trial data are credible.
- **Good Manufacturing Practice (GMP):** A system for ensuring that products are consistently produced and controlled according to quality standards. It covers all aspects of production from the raw materials, premises, and equipment to the training and personal hygiene of staff.

8.9 Regulatory Pathways for New Drug Approval

While the CTD format is standard, the procedural pathways through which a drug is approved vary by region, often reflecting different legal frameworks and public health priorities.

8.9.1 United States

The U.S. system is characterized by the **Investigational New Drug (IND)** application, which must be filed before a sponsor can begin clinical trials in humans. Once trials are complete, the **New Drug Application (NDA)** (for small molecules) or **Biologics License Application (BLA)** (for biologics) is submitted for marketing approval. For generic drugs, the **Abbreviated New Drug Application (ANDA)** is used, which does not require preclinical or clinical data to establish safety and efficacy but must demonstrate that the generic version is bioequivalent to the innovator drug.

8.9.2 Europe

The EU offers several routes for marketing authorization:

- **Centralized Procedure:** The EMA evaluates the application, and the European Commission grants a single authorization valid in all EU/EEA member states. This is mandatory for medicines derived from biotechnology and for those intended to treat HIV/AIDS, cancer, diabetes, and rare diseases.
- **Decentralized Procedure (DCP):** Used for medicines not yet authorized in any EU country and that do not fall under the mandatory centralized route. The sponsor applies for authorization in several member states simultaneously.
- **Mutual Recognition Procedure (MRP):** Used when a medicine is already authorized in one EU member state. The sponsor seeks recognition of that authorization in other member states.

BIOPRESS

8.9.3 India

The approval process in India is governed by the NDCT Rules, 2019. Sponsors must obtain clinical trial approval from the DCGI. Following successful trials, a New Drug Application is submitted. The CDSCO emphasizes the role of the SECs in providing expert scientific assessment. Recent reforms have significantly reduced the time required for clinical trial and new drug approvals to foster a more favorable environment for innovation.

8.10 Accelerated and Expedited Approval Pathways

Recognizing that patients with life-threatening illnesses cannot always wait for traditional approval timelines, agencies have developed expedited pathways:

- **Fast Track (USA):** Designed to facilitate the development and expedite the review of drugs to treat serious conditions and fill an unmet medical need.
- **Breakthrough Therapy (USA):** Granted for drugs that demonstrate substantial improvement over existing therapies on one or more clinically significant endpoints.
- **Sakigake (Japan):** A strategy to introduce innovative medicines to the Japanese market first. It provides for a shortened review period (target 6 months) and priority consultations for drugs that are novel and address serious conditions.
- **PRIME (Priority Medicines, Europe):** A scheme that provides enhanced support for the development of medicines that target an unmet medical need, aiming to bring them to patients earlier.
- **Conditional Marketing Authorization (Europe/India):** Granted for medicines where the benefit of immediate availability outweighs the risk of less-than-complete data, typically for urgent public health needs.

Table 8.3: Different types of pathways of different countries

Pathway	Region	Focus	Benefit
Fast Track	USA	Serious conditions	Frequent meetings, rolling review.[1]

BIOPRESS

Sakigake	Japan	First-in-world innovation	6-month review target.[3]
PRIME	Europe	Major public health interest	Early interaction, accelerated assessment.[1]
Priority Review	USA/India	Significant improvement	6-month vs. 10-month review (US).[1]

8.11 Clinical Trial Regulatory Considerations

The regulation of clinical trials is the most critical juncture of drug development, as it involves human exposure to experimental agents.

8.11.1 Ethical Approvals and Informed Consent

Every clinical trial must be reviewed and approved by an **Institutional Review Board (IRB)** or **Independent Ethics Committee (IEC)** to ensure that the risks to participants are minimized and justified by the potential benefits.[18] **Informed Consent** is a non-negotiable ethical requirement; participants must be fully informed of the trial's nature, risks, and benefits and must voluntarily agree to participate without coercion.

8.11.2 Trial Registration and Data Integrity

Transparency is enforced through mandatory trial registration on public databases like ClinicalTrials.gov. This ensures that the global medical community is aware of ongoing research and prevents the selective publication of only positive results. Data integrity is maintained through rigorous monitoring, auditing, and adherence to the protocol. The emergence of electronic systems has made data integrity more manageable but has also introduced new cybersecurity and privacy risks that regulators must address.

8.12 Chemistry, Manufacturing, and Controls (CMC)

The CMC section of a regulatory submission is often where technical hurdles lead to approval delays. It encompasses the documentation required to prove that the drug can be manufactured consistently to a high standard.

8.12.1 Drug Substance and Product Quality

CMC covers the complete characterization of the drug substance, including its synthesis, purification, and physical properties. For the drug product, it includes the formulation, packaging, and the validation of the manufacturing process.[21] Under ICH Q8, sponsors use **Critical Quality Attributes (CQAs)** and **Critical Process Parameters (CPPs)** to define a "Design Space," which allows for manufacturing flexibility without requiring regulatory re-approval for minor changes.

8.12.2 Stability Studies

Stability testing is essential for determining a drug's shelf-life and storage conditions. The ICH Q1 series provides a harmonized framework for these studies.

- **ICH Q1A(R2):** Outlines the standard protocol for stability testing under different climatic zones.
- **Accelerated Testing:** $40^\circ\text{C} \pm 2^\circ\text{C}/75\%\ \text{RH} \pm 5\%\ \text{RH}$ for 6 months.
- **Long-term Testing:** $25^\circ\text{C} \pm 2^\circ\text{C}/60\%\ \text{RH} \pm 5\%\ \text{RH}$ for 12 months (Zone II).

The 2025/2026 consolidation of the ICH Q1 guidelines aims to modernize these standards, including specific considerations for advanced therapy medicinal products (ATMPs) and vaccines.

8.12.3 Scale-up and Manufacturing Compliance

Moving from laboratory-scale batches used in clinical trials to commercial-scale manufacturing is a significant challenge. Any changes in the manufacturing process post-approval must be validated to ensure they do not impact the drug's safety or efficacy. This is particularly critical for biologics, where minor changes in fermentation or purification can significantly alter the final protein product.

8.13 Regulatory Submission Dossiers

The format of regulatory submissions is now almost entirely standardized through the Common Technical Document (CTD), developed by the ICH.

8.13.1 Common Technical Document (CTD) Modules

The CTD is organized into five modules:

- **Module 1 (Regional Administrative Information):** Contains documents such as the application form, proposed labeling, and cover letter. This module is specific to the region where the application is submitted.
- **Module 2 (CTD Summaries):** High-level overviews and summaries of the data in Modules 3, 4, and 5. This includes the **Quality Overall Summary (QOS)**, which should ideally not exceed 40-80 pages.
- **Module 3 (Quality):** The CMC section, providing detailed pharmaceutical documentation on the drug substance and product.
- **Module 4 (Non-clinical Study Reports):** Contains the full reports of pharmacology and toxicology studies conducted in the laboratory and in animals.
- **Module 5 (Clinical Study Reports):** The most extensive module, containing the full reports of all clinical trials, including pharmacokinetic, efficacy, and safety analyses.

8.13.2 Electronic CTD (eCTD)

The eCTD is the electronic version of the CTD and is now mandatory for most submissions to the FDA, EMA, and PMDA. It uses an XML backbone to manage the life-cycle of the submission, allowing for easy updates and cross-referencing. The transition to eCTD v4.0 is currently underway globally, promising even greater efficiency and integration.

8.14 Pharmacovigilance and Post-Marketing Surveillance

The regulatory process does not end with a drug's approval. Post-marketing surveillance is critical for detecting rare adverse events that may only become apparent when the drug is used by a larger, more diverse population over a longer period.

8.14.1 Periodic Safety Reports

Sponsors are required to submit regular safety reports to regulators:

- **Periodic Safety Update Report (PSUR):** Provides a summary of the safety profile of a medicinal product during a specific reporting period.
- **Periodic Benefit-Risk Evaluation Report (PBRER):** A more advanced format introduced by ICH E2C(R2) that evaluates both the benefits and risks of a drug to provide a comprehensive scientific analysis of its total clinical value.[27]

BIOPRESS

- **Development Safety Update Report (DSUR):** Focuses on the safety of trial participants during the investigational phase of a drug's development.[23]

8.14.2 Risk Management Plans (RMP)

An RMP is a dynamic document that describes the known and potential risks of a drug and the measures taken to minimize those risks. This may include additional safety studies, educational programs for healthcare providers, or restricted distribution systems (e.g., REMS in the U.S.).

Table 8.4: Various types of features and their stages

Feature	**PSUR**	**PBRER**	**DSUR**
Phase	Post-approval	Post-approval	Pre-approval (Clinical)
Focus	Safety updates	Benefit-risk balance	Trial participant safety
Requirement	Cumulative safety data	Integrated analysis	Annual updates
Guideline	Regional	ICH E2C(R2)	ICH E2F

BIOPRESS

8.15 Challenges in Drug Approval

Despite harmonization, the drug approval process faces several systemic challenges.

8.15.1 Regulatory Delays and Global Variation

Differences in clinical data requirements, such as the historical need for Japanese-specific Phase I data, have led to significant "drug lags" where Japanese patients waited years longer than their U.S. or European counterparts for new therapies. While reforms are narrowing these gaps, variations in agency interpretation of efficacy data can still lead to discordant decisions.

8.15.2 High Cost of Compliance and Data Transparency

The immense cost of meeting regulatory standards—estimated to be a significant portion of the $1-$2 billion development cost per drug—can stifle innovation, particularly for small biotech companies. Furthermore, the push for greater data transparency must be balanced against the need to protect commercial intellectual property and patient privacy.

8.15.3 Ethical and Legal Considerations

Ethical considerations are woven into every stage of drug development. Beyond clinical trial ethics, regulators must address data privacy in the era of digital health. The use of patient data from electronic health records (EHRs) for regulatory purposes requires robust legal frameworks to prevent misuse.[40] Legal liabilities for manufacturers remain a constant concern, particularly when post-marketing data reveals previously unknown risks, leading to potential litigation and product withdrawals.

8.16 Emerging Trends in Regulatory Science

The most profound shifts in drug regulation are being driven by technology and the paradigm of precision medicine.

8.16.1 Artificial Intelligence (AI) and Machine Learning (ML)

AI is being integrated across the drug lifecycle, from identifying "undruggable" targets via Targeted Protein Degradation (TPD) to optimizing leads.[14] In 2024, AI's role was cemented by the discovery of molecules like **Halicin** (a novel antibiotic) and the progression of **ISM001-055** (Rentosertib) for idiopathic pulmonary fibrosis into Phase II clinical trials. AI-driven analysis of multi-omics data is also refining the identification of biomarkers, which are crucial for stratifying patients in precision oncology.

8.16.2 Real-World Evidence (RWE)

RWE—evidence derived from data collected outside of traditional RCTs—is gaining regulatory acceptance. The FDA’s 2018 Framework and subsequent 2023-2025 case studies (e.g., **Fosdenopterin**, **Vosoritide**) demonstrate how RWE can provide confirmatory evidence for efficacy or act as a historical control in rare disease trials where RCTs are not feasible.

8.16.3 Multi-Omics Integration

The integration of genomics, proteomics, and metabolomics is enabling a more granular understanding of disease biology. Regulators are increasingly evaluating "multi-modal" signatures as surrogate endpoints or as diagnostic tools (CDx) to guide treatment.

8.17 Future Perspectives

The future of drug regulation will likely be characterized by **Digital Transformation**. The adoption of eCTD v4.0 and the potential for "rolling reviews" of real-time data will further shorten approval timelines. We also anticipate the rise of **Global Harmonization through Reliance**, where regulatory agencies in smaller or emerging markets increasingly rely on the scientific assessments of mature authorities like the FDA or EMA to grant local approvals, thereby improving global patient access.

Personalized medicine will necessitate even more flexible regulations. As therapies move toward "N-of-1" designs (e.g., personalized cancer vaccines), the traditional batch-release models of GMP will need to evolve. The use of AI in regulatory decision-making—potentially through automated triage of safety reports or AI-assisted review of CMC data—is another frontier that is currently being explored.

8.18 Conclusion

The regulatory frameworks governing new drug approval are the essential safeguard of the modern pharmaceutical industry. While the technical requirements for safety, efficacy, and quality are becoming increasingly complex, the tools available to meet those standards are also advancing. The harmonization efforts of the ICH and the emergence of expedited pathways have significantly improved patient access to innovative therapies. However, the future will demand even greater agility from regulators and sponsors alike. Embracing AI, RWE, and digital transformation while maintaining the bedrock principles of medical ethics and rigorous scientific scrutiny will be paramount. As we move further into the era of precision medicine, the synergy between regulatory innovation and scientific discovery will determine our ability to combat global health challenges effectively and safely.

8.19 References

1. Global Regulatory Frameworks for Drug Approval: A Comparative ..., https://www.ijdra.com/index.php/journal/article/download/852/437
2. Global Regulatory Frameworks for Drug Approval: A Comparative Review of USFDA, EMA, and CDSCO, https://www.ijdra.com/index.php/journal/article/view/852
3. Japan's PMDA Drug Approval: Standard & Expedited Pathways - IntuitionLabs, https://intuitionlabs.ai/articles/pmda-regulatory-pathways-japan
4. Japan's Drug Approval Process: A Guide to PMDA & MHLW | IntuitionLabs, https://intuitionlabs.ai/articles/japan-drug-approval-process

5. Review of AI/ML applications in medicines lifecycle (2024) - EMA, https://www.ema.europa.eu/en/documents/report/review-artificial-intelligence-machine-learning-applications-medicines-lifecycle-2024-horizon-scanning-short-report_en.pdf

6. Artificial Intelligence in Small-Molecule Drug Discovery: A Critical Review of Methods, Applications, and Real-World Outcomes - MDPI, https://www.mdpi.com/1424-8247/18/9/1271

7. Developing a Comprehensive Strategy for Implementing AI & Multi-Omics for Translational Research - Precision for Medicine, https://www.precisionformedicine.com/blog/developing-a-comprehensive-strategy-for-implementing-ai-multi-omics-for-translational-research

8. Multi-omics strategies for biomarker discovery and application in personalized oncology, https://pubmed.ncbi.nlm.nih.gov/41269529/

9. FDA use of Real-World Evidence in Regulatory Decision Making, https://www.fda.gov/science-research/real-world-evidence/fda-use-real-world-evidence-regulatory-decision-making

10. FDA publishes new landscape analysis of real-world evidence in regulatory decision-making, https://becarispublishing.com/digital-content/blog-post/fda-publishes-new-landscape-analysis-real-world-evidence-regulatory-decision-making

11. Expediting Drug Development in Japan: A PMDA Perspective - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12641066/

12. ICH guidelines (Q,S,E,M) | PPTX - Slideshare, https://www.slideshare.net/slideshow/ich-guidelines-qsem/223429941

13. ICH Quality Guidelines Summary Table | PDF - Scribd, https://www.scribd.com/document/895923112/Summary

14. New strategies to enhance the efficiency and precision of drug ..., https://pmc.ncbi.nlm.nih.gov/articles/PMC11850385/

15. Lead Optimization in Drug Discovery | Danaher Life Sciences, https://lifesciences.danaher.com/us/en/library/lead-optimization-drug-discovery-guide.html

16. Progress of AI-Driven Drug–Target Interaction Prediction and Lead Optimization - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC12563295/

17. Drug Discovery → Lead Optimization - Developing Medicines, https://drugdevelopment.web.unc.edu/drug-discovery-%E2%86%92-lead-optimization/

18. (PDF) Ich Guidelines - ResearchGate, https://www.researchgate.net/publication/398232285_Ich_Guidelines

19. 2024 New Drug Therapy Approvals Annual Report - FDA, https://www.fda.gov/files/drugs/published/new-drug-therapy-2025-annual-report.pdf

BIOPRESS

20. ICOTYDE™ (icotrokinra) one-year results confirm lasting skin clearance and favorable safety profile in once-daily pill for plaque psoriasis, https://www.jnj.com/media-center/press-releases/icotyde-icotrokinra-one-year-results-confirm-lasting-skin-clearance-and-favorable-safety-profile-in-once-daily-pill-for-plaque-psoriasis

21. Regulatory Writing Modules 1-5: Submit CTD successfully - Aristo Group, https://www.aristo-group.com/en/guide/regulatory-writing-modules-1-5-submit-ctd-successfully/

22. An overview of the Common Technical Document (CTD) regulatory dossier, https://journal.emwa.org/regulatory-writing-basics/an-overview-of-the-common-technical-document-ctd-regulatory-dossier/article/1693/2047480614z2e000000000207.pdf

23. Categorizing & Managing Risks in Aggregate Safety Reports - Certara, https://www.certara.com/blog/categorizing-managing-risks-in-aggregate-safety-reports/

24. Comparison of drug approvals of the FDA and EMA between 2013 and 2023 - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12894151/

25. Guide to ICH Q7, Q8, & Q9: GMP, QbD, and QRM Standards | IntuitionLabs, https://intuitionlabs.ai/articles/ich-q7-q8-q9-guide

26. ICH Official web site : ICH, https://www.ich.org/

27. PBRER vs PSUR in Pharmacovigilance: Key Differences Guide - Atvigilx, https://atvigilx.com/pbrer-vs-psur-key-differences-explained/

28. 3. Format of the Submission – the Common Technical Document (CTD), https://learning.eupati.eu/mod/book/tool/print/index.php?id=904

29. Real-World Evidence in FDA Approvals for Labeling Extension of Drugs and Biologics,https://www.ispor.org/heor-resources/presentations-database/presentation-cti/ispor-2025/poster-session-2/real-world-evidence-in-fda-approvals-for-labeling-extension-of-drugs-and-biologics

30. Common Technical Document (CTD) Structure & Module Overview - Pubrica, https://pubrica.com/services/medical-writing/regulatory-writing/ctd-structure-overview-modules/

31. Molecular Biomarkers in Neurological Diseases: Advances in Diagnosis and Prognosis, https://www.mdpi.com/1422-0067/26/5/2231

32. Multi-Omics Integration in Personalized Medicine: Advancing Laboratory Diagnostics and Precision Therapeutics in the Era of Individualized Healthcare, https://jddtonline.info/index.php/jddt/article/view/7112

33. Common Technical Document (CTD) Preparation - ScienceDocs, https://www.sciencedocs.com/common-technical-document/

BIOPRESS

34. ICH Q1 Stability Testing Guidelines - MasterControl, https://www.mastercontrol.com/glossary-page/ich-q1-stability-testing-guidelines/

35. Q1 Stability Testing of Drug Substances and Drug Products - FDA, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/q1-stability-testing-drug-substances-and-drug-products

36. ICH Q1: Stability Testing of Drug Substances and Drug Products Step 2 document – to be released for comments 11 April 2025, https://database.ich.org/sites/default/files/ICH_Q1_Step2_Presentation_2025_0411_0.pdf

37. DSUR vs. PSUR/PBRER: A Pharmacovigilance Comparison | IntuitionLabs, https://intuitionlabs.ai/articles/dsur-vs-psur-pbrer

38. PSUR vs. PBRER: Key Differences Explained – pharmacovigilance research, https://www.pharmaregulatory.in/psur-vs-pbrer-key-differences-explained-pharmacovigilance-research/

39. Preparation of pharmacovigilance regulatory documents: PSURs/PBRERs, RMPs, DSURs, ACO - NextPV Services, https://www.nextpvservices.com/safety-corner/preparation-of-pharmacovigilance-regulatory-documents-psursp/

40. Challenges and potential of using digital biomarkers in healthcare and clinical trials - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12996552/

41. Top 10 Most Popular Drug Hunter Case Studies of 2025, https://drughunter.com/articles/top-10-most-popular-drug-hunter-case-studies-of-2025

42. Precision medicine meets machine learning: AI and oncology biomarkers | pharmaphorum, https://pharmaphorum.com/digital/precision-medicine-meets-machine-learning-ai-and-oncology-biomarkers

43. Lost in Translation: Turning Multi-omic Chaos into Clinical Clarity - Precision for Medicine, https://www.precisionformedicine.com/blog/lost-in-translation-turning-multi-omics-chaos-into-clinical-clarity

BIOPRESS

TITLE - CONTEMPORARY DRUG DESIGN & THERAPEUTIC INNOVATION

Modern drug discovery is being transformed by artificial intelligence, computational biology, molecular modeling, precision therapeutics, and translational research. Contemporary Drug Design & Therapeutic Innovation presents a comprehensive overview of emerging approaches that are reshaping pharmaceutical and biotechnology research.

This book covers key areas including:

- Artificial Intelligence in Drug Delivery
- Structure-Based Drug Design & Molecular Docking
- Lead Optimization Strategies
- ADMET Prediction & Pharmacokinetic Modeling
- Drug Repurposing & Computational Screening
- Biomarkers in Precision Therapeutics
- Translational Cancer Research
- Regulatory Pathways in Drug Approval

Designed for students, researchers, academicians, and healthcare professionals, this book bridges foundational concepts with modern therapeutic innovations and practical applications in pharmaceutical sciences.

Contributing Authors

Dr Richa Sharma, Gajanand Sharma, H.M Rakshith, Jagruti Gosavi, Rajeshwari Dughde, Trupti Kadam, Ashwini Badhe, Dr K. P. Beena, Arghya Saha Chowdhury, Ilias Ahmed, Paragmoni Bora, Dr C. Jaganmohan, Dr Srinidhi S., Dr Sasikala, E. Nivetha, and R. Geerhanjali.

Editor

Dr. Richa Sharma

Professor, Department of Microbiology,

Mahatma Gandhi University of Medical Sciences & Technology, Jaipur, India

Co-Editor

Dr. K. P. Beena

Associate Professor, Department of Pharmaceutical Chemistry,

Sri Ramakrishna Institute of Paramedical Sciences, College of Pharmacy,

Coimbatore, Tamil Nadu, India

Published by

Biopress Division, Swalife Foundation

Email: swalifefoundation@gmail.com

Website: www.biopresspublication.com

MRP: ₹550/-

(INCLUSIVE OF ALL TAXES)

9 789349 682740